

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
26 May 2005 (26.05.2005)

PCT

(10) International Publication Number  
WO 2005/047850 A2

(51) International Patent Classification<sup>7</sup>: G01N  
(21) International Application Number: PCT/US2004/032753  
(22) International Filing Date: 4 October 2004 (04.10.2004)  
(25) Filing Language: English  
(26) Publication Language: English  
(30) Priority Data: 10/677,877 2 October 2003 (02.10.2003) US  
(71) Applicant: GENHUNTER CORPORATION [US/US]; 624 Grassmere Park Drive, Suite 17, Nashville, TN 37211 (US).  
(72) Inventor: LIANG, Peng; 9 Castlewood Court, Nashville, TN 37215 (US).  
(74) Agent: ARCHIBALD, Patrick, D.; Lambert & Associates, 92 State Street, Boston, MA 02109 (US).

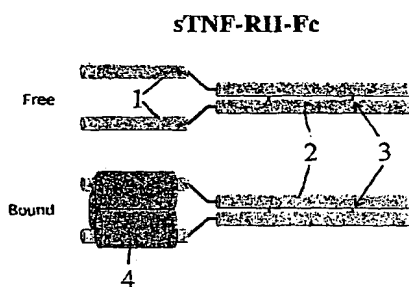
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

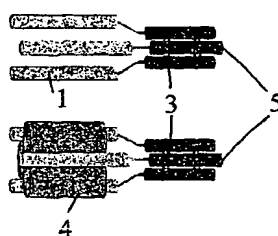
(54) Title: METHODS AND COMPOSITIONS FOR PRODUCING SECRETED TRIMERIC RECEPTOR ANALOGS AND BIOLOGICALLY ACTIVE FUSION PROTEINS

Side-View



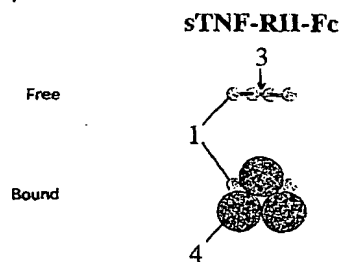
A

sTNF-RII-Trimer



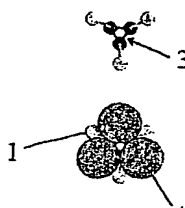
C

Top-View



B

sTNF-RII-Trimer



D

(57) Abstract: Methods and compositions for producing secreted soluble receptors and biologically active polypeptides in trimeric forms are disclosed. The process involves fusing the DNA template encoding a soluble receptor with a ligand binding domain or biologically active polypeptide to a DNA sequence encoding a C-propeptide of collagen, which is capable of self-assembly into a covalently linked trimer. The resulting fusion proteins are secreted as trimeric soluble receptor analogs, which can be used for more efficient neutralization of the biological activities of their naturally occurring trimeric ligands.

BEST AVAILABLE COPY

**Declarations under Rule 4.17:**

- as to the identity of the inventor (Rule 4.17(i)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS,

JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations

**Published:**

- without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

## METHODS AND COMPOSITIONS FOR PRODUCING SECRETED TRIMERIC RECEPTOR ANALOGS AND BIOLOGICALLY ACTIVE FUSION PROTEINS

### 5 FIELD OF THE INVENTION

The present invention relates to methods for protein expression, and more specifically, for creating and expressing secreted and biologically active trimeric proteins, such as trimeric soluble receptors.

### 10 BACKGROUND OF INVENTION

In multicellular organisms, such as humans, cells communicate with each other by the so-called signal transduction pathway, in which a secreted ligand (e.g. cytokines, growth factors or hormones) binds to its cell surface receptor(s), leading to receptor activation. The receptors are membrane proteins, which consist of an extracellular domain responsible for ligand binding, a central transmembrane region followed by a cytoplasmic domain responsible for sending the signal downstream. Signal transduction can take place in the following three ways: paracrine (communication between neighboring cells), autocrine (cell communication to itself) and endocrine (communication between distant cells through circulation), depending on the source of a secreted signal and the location of target cell expressing a receptor(s). One of the general mechanisms underlying receptor activation, which sets off a cascade of events beneath the cell membrane including the activation of gene expression, is that a polypeptide ligand such as a cytokine, is present in an oligomeric form, such as a homo-dimer or trimer, which when bound to its monomeric receptor at the cell outer surface, leads to the oligomerization of the receptor. Signal transduction pathways play a key role in normal cell development and differentiation, as well as in response to external insults such as bacterial and viral

infections. Abnormalities in such signal transduction pathways, in the form of either underactivation (e.g. lack of ligand) or overactivation (e.g. too much ligand), are the underlying causes for pathological conditions and diseases such as arthritis, cancer, AIDS, and diabetes.

5           One of the current strategies for treating these debilitating diseases involves the use of receptor decoys, such as soluble receptors consisting of only the extracellular ligand-binding domain, to intercept a ligand and thus overcome the overactivation of a receptor. The best example of this strategy is the creation of Enbrel, a dimeric soluble TNF- $\alpha$  receptor-immunoglobulin (IgG) fusion protein by  
10 Immunex (Mohler et al., 1993; Jacobs et al., 1997), which is now part of Amgen. The TNF family of cytokines is one of the major pro-inflammatory signals produced by the body in response to infection or tissue injury. However, abnormal production of these cytokines, for example, in the absence of infection or tissue injury, has been shown to be one of the underlying causes for diseases such as arthritis and psoriasis.  
15 Naturally, a TNF- $\alpha$  receptor is present in monomeric form on the cell surface before binding to its ligand, TNF- $\alpha$ , which exists, in contrast, as a homotrimer (Locksley et al., 2001). Accordingly, fusing a soluble TNF- $\alpha$  receptor with the Fc region of immunoglobulin G1, which is capable of spontaneous dimerization via disulfide bonds (Sledziewski et al., 1992 and 1998), allowed the secretion of a dimeric soluble  
20 TNF- $\alpha$  receptor (Mohler et al., 1993; Jacobs et al., 1997). In comparison with the monomeric soluble receptor, the dimeric TNF- $\alpha$  receptor II -Fc fusion has a greatly increased affinity to the homo-trimeric ligand. This provides a molecular basis for its clinical use in treating rheumatoid arthritis (RA), an autoimmune disease in which constitutively elevated TNF- $\alpha$ , a major pro-inflammatory cytokine, plays an  
25 important causal role. Although Enbrel was shown to have a  $K_i$  in the pM range

(ng/mL) to TNF- $\alpha$  ( Mohler et al., 1993), 25 mg twice a week subcutaneous injections, which translates to  $\mu$ g/mL level of the soluble receptor, are required for the RA patients to achieve clinical benefits (www.enbrel.com). The high level of recurrent Enbrel consumption per RA patients has created a great pressure as well as  
5 high cost for the drug supply, which limits the accessibility of the drug to millions of potential patients in this country alone.

In addition to the TNF- $\alpha$  family of potent proinflammatory cytokines, the HIV virus that causes AIDS also uses a homo-trimeric coat protein, gp120, to gain entry into CD-4 positive T helper cells in our body (Kwong et al., 1998). One of the  
10 earliest events during HIV infection involves the binding of gp120 to its receptor CD-4, uniquely expressed on the cell surface of T helper cells (Clapham et al., 2001). Monomeric soluble CD-4 was shown over a decade ago as a potent agent against HIV infection (Clapham et al., 1989) however, the excitement was sadly dashed when its potency was shown to be limited only to laboratory HIV isolates (Daar et al., 1990).  
15 It turned out that HIV strains from AIDS patients, unlike the laboratory isolates, had a much lower affinity to the monomeric soluble CD-4, likely due to the sequence variation on the gp120 (Daar et al., 1990). Although the dimeric soluble CD-4-Fc fusion proteins have been made, these decoy CD-4 HIV receptors showed little antiviral effect against natural occurring HIVs from AIDS patients, both in the  
20 laboratories and in clinics, due to the low affinity to the gp120 (Daar et al., 1990).

Clearly, there is a great need to be able to create secreted homo-trimeric soluble receptors or biologically active proteins, which can have perfectly docked binding sites, hence higher affinity, to their naturally occurring homo-trimeric ligands, such as the TNF family of cytokines and HIV coat proteins. Such trimeric receptor  
25 decoys theoretically should have a much higher affinity than its dimeric counterparts

to their trimeric ligand. Such rationally designed soluble trimeric receptor analogs could significantly increase the clinical benefits as well as lower the amount or frequency of the drug injections for each patient. To be therapeutically feasible, like immunoglobulin Fc, the desired trimerizing protein moiety should ideally be part of a naturally secreted protein that is both abundant in the body and capable of efficient self-trimerization.

Collagen is a family of fibrous proteins that are the major components of the extracellular matrix. It is the most abundant protein in mammals, constituting nearly 25% of the total protein in the body. Collagen plays a major structural role in the formation of bone, tendon, skin, cornea, cartilage, blood vessels, and teeth (Stryer, 1988). The fibrillar types of collagen I, II, III, IV, V, and XI are all synthesized as larger trimeric precursors, called procollagens, in which the central uninterrupted triple-helical domain consisting of hundreds of "G-X-Y" repeats (or glycine repeats) is flanked by non-collagenous domains (NC), the N- propeptide and the C-propeptide (Stryer, 1988). Both the C- and N-terminal extensions are processed proteolytically upon secretion of the procollagen, an event that triggers the assembly of the mature protein into collagen fibrils which forms an insoluble cell matrix (Prockop et al., 1998). The shed trimeric C-propeptide of type I collagen is found in the blood of normal people at a concentration in the range of 100-600 ng/mL, with children having a higher level which is indicative with active bone formation.

Type I, IV, V and XI collagens are mainly assembled into heterotrimeric forms consisting of either two  $\alpha$ -1 chains and one  $\alpha$ -2 chain (for Type I, IV, V), or three different  $\alpha$  chains (for Type XI), which are highly homologous in sequence. The type II and III collagens are both homotrimers of  $\alpha$ -1 chain. For type I collagen, the most abundant form of collagen, stable  $\alpha$ -1(I) homotrimer is also formed and is

present at variable levels (Alvarès et al., 1999) in different tissues. Most of these collagen C-propeptide chains can self-assemble into homotrimers, when over-expressed alone in a cell. Although the N-propeptide domains are synthesized first, molecular assembly into trimeric collagen begins with the in-register association of the C-propeptides. It is believed the C-propeptide complex is stabilized by the formation of interchain disulfide bonds, but the necessity of disulfide bond formation for proper chain registration is not clear. The triple helix of the glycine repeats and is then propagated from the associated C-termini to the N-termini in a zipper-like manner. This knowledge has led to the creation of non-natural types of collagen matrix by swapping the C-propeptides of different collagen chains using recombinant DNA technology (Bulleid et al., 2001). Non-collagenous proteins, such as cytokines and growth factors, also have been fused to the N-termini of either pro-collagens or mature collagens to allow new collagen matrix formation, which is intended to allow slow release of the noncollagenous proteins from the cell matrix (Tomita et al., 2001). However, under both circumstances, the C-propeptides are required to be cleaved before recombinant collagen fibril assembly into an insoluble cell matrix.

#### SUMMARY OF THE INVENTION

Disclosed here is an invention that allows any soluble receptors or biologically active polypeptides to be made into trimeric forms as secreted proteins. The essence of the invention is to fuse any soluble receptors and biologically active proteins in-frame to the C-propeptide domain of fibrillar collagen, which is capable of self-trimerization, using recombinant DNA technology. The resulting fusion proteins when expressed in eukaryotic cells are secreted as soluble proteins essentially all in

trimeric forms covalently strengthened by inter-molecular disulfide bonds formed among three C-propeptides.

In one aspect of the invention, a method for producing secreted trimeric fusion proteins is disclosed, comprising the following: (a) introducing into a eukaryotic host  
5 cell a DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be trimerized, which in turn is joined in-frame to the C-terminal portion of collagen capable of self-trimerization; (b) growing the host cell in an appropriate growth medium under physiological conditions to allow the secretion  
10 of a trimerized fusion protein encoded by said DNA sequence; and (c) isolating the secreted trimeric fusion protein from a host cell.

Within one embodiment, the signal peptide sequence is the native sequence of the protein to be trimerized. Within another embodiment, the signal peptide sequence is from a secreted protein different from that to be trimerized. Within one  
15 embodiment, the non-collagen polypeptide to be trimerized is a soluble receptor consisting of the ligand binding domain(s). Within one embodiment, the C-terminal portion of collagen is the C-propeptide without any triple helical region of collagen (Sequence IDs: 3-4). Within another embodiment, the C-terminal collagen consists of a portion of the triple helical region of collagen as linker to the non-collagenous  
20 proteins to be trimerized (Sequence IDs: 1-2). Within another embodiment, the C-terminal portion of collagen has a mutated or deleted BMP-1 protease recognition site (Sequence IDs: 3-4).

In one aspect of the invention, a method for producing a secreted trimeric fusion protein is disclosed, comprising the following: (a) introducing into a eukaryotic  
25 host cell a DNA construct comprising a promoter which drives the transcription of an



open reading frame consisting of a signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be trimerized, which in turn is joined in-frame to the C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V),  
5 pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (b) growing the host cell in an appropriate growth medium under physiological conditions to allow the secretion of a trimerized fusion protein encoded by said DNA sequence; and (c) isolating the secreted trimeric fusion protein from a host cell.

In a preferred embodiment, the non-collagen polypeptide to be trimerized is  
10 the soluble TNF-RII (p75) (Sequence IDs: 9-12). In another preferred embodiment, the non-collagen polypeptide to be trimerized is soluble CD-4, the co-receptor of HIV (Sequence IDs: 13-16). In yet another preferred embodiment, the non-collagen polypeptide to be trimerized is a placental secreted alkaline phosphatase (Sequence IDs: 5-8).

15 In one aspect of the invention, a method for producing a secreted trimeric fusion protein is disclosed, comprising the following: (a) introducing into a eukaryotic host cell a first DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be trimerized, which in turn is joined in-frame  
20 to the C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (b) introducing into a eukaryotic host cell a second DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a second signal peptide  
25 sequence which is linked in-frame to a second non-collagen polypeptide to be

trimerized, which in turn is joined in-frame to the second C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (c) growing the host cell in an appropriate  
5 growth medium under physiological conditions to allow the secretion of a trimerized fusion protein encoded by said first and second DNA sequences; and (d) isolating the secreted trimeric fusion protein from the host cell.

In one aspect of the invention, a method for producing a secreted trimeric fusion protein is disclosed, comprising the following: (a) introducing into a eukaryotic  
10 host cell a first DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be trimerized, which in turn is joined in-frame to the C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V),  
15 pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (b) introducing into a eukaryotic host cell a second DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a second signal peptide sequence which is linked in-frame to a second non-collagen polypeptide to be trimerized, which in turn is joined in-frame to a second C-terminal portion of collagen  
20 capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (c) introducing into a eukaryotic host cell a third DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a third signal peptide sequence which is linked in-frame  
25 to a third non-collagen polypeptide to be trimerized, which in turn is joined in-frame

to a third C-terminal portion of collagen capable of self-trimerization; selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (d) growing the host cell in an appropriate growth medium under physiological conditions to allow the  
5 secretion of a trimerized fusion protein encoded by said first and second DNA sequences; and (e) isolating the secreted trimeric fusion protein from the host cell.

The following are the advantages of this invention: (1) collagen is the most abundant protein secreted in the body of a mammal, constituting nearly 25% of the total proteins in the body; (2) the major forms of collagen naturally occur as trimeric  
10 helices, with their globular C-propeptides being responsible for the initiating of trimerization; (3) the trimeric C-propeptide of collagen proteolytically released from the mature collagen is found naturally at sub microgram/mL level in the blood of mammals and is not known to be toxic to the body; (4) the linear triple helical region of collagen can be included as a linker with predicted 2.9 Å spacing per residue, or  
15 excluded as part of the fusion protein so the distance between a protein to be trimerized and the C-propeptide of collagen can be precisely adjusted to achieve an optimal biological activity; (5) the recognition site of BMP1 which cleaves the C-propeptide off the pro-collagen can be mutated or deleted to prevent the disruption of a trimeric fusion protein; (6) the C-proptide domain provides a universal affinity tag,  
20 which can be used for purification of any secreted fusion proteins created by this invention.

In contrast to the Fc Tag technology (Sledziewski et al., 1992 and 1998), with which secreted dimeric fusion proteins can be created, this timely invention disclosed herein enables the creation and secretion of soluble trimeric fusion proteins for the  
25 first time. Given the fact that a homotrimer has 3-fold symmetry, whereas a

homodimer has only 2-fold symmetry, the two distinct structural forms theoretically can never be perfectly overlaid (Fig 1). As such, neither the homodimeric soluble TNF-R-Fc (e.g. Enbrel), nor the soluble CD4-Fc fusion proteins, could have had an optimal interface for binding to their corresponding homotrimeric ligands, TNF- $\alpha$  and HIV gp120, respectively. In contrast, homotrimeric soluble TNF receptors and CD4 created by the current invention are trivalent and structurally have the potential to perfectly dock to the corresponding homotrimeric ligands. Thus, these trimeric soluble receptor analogs can be much more effective in neutralizing the biological activities of their trimeric ligands. With this timely invention, more effective yet less expensive drugs, such as trimeric soluble TNF-R and CD4 described in the preferred embodiments, can be readily and rationally designed to combat debilitating diseases such as arthritis and AIDS. Trimeric soluble gp120 can also be created with this invention, which could better mimic the native trimeric gp120 coat protein complex found on HIV viruses, and used as a more effective vaccine compared to non-trimeric gp120 antigens previously used. Also chimeric antibodies in trimeric form can be created with the current invention, which could endow greatly increased avidity of an antibody in neutralizing its antigen.

#### BRIEF DESCRIPTION OF DRAWINGS

Fig.1A, Fig. 1B, Fig. 1C and Fig. 1D is a schematic representation of the method according to the invention compared to prior dimeric immunoglobulin Fc fusion. Fig 1A is a side elevation view and Fig. 1B is a top plan view: Structural characteristics of a homodimeric soluble sTNF RII receptor-Fc fusion, such as Amgen's Enbrel, in either ligand-free or -bound form as indicated.

Domains labeled 1 denote soluble TNF-RII. Note that the Fc (labeled as 2 with inter-chain disulfide bonds 3) fusion protein is dimeric in structure. Given its 2-fold symmetry, the dimeric Fc fusion protein is bivalent and thus theoretically does not have the optimal conformation to bind to a homotrimeric ligand, such as TNF- $\alpha$  (labeled 4), which has a 3-fold symmetry.

Fig. 1C is a side elevation view and Fig. 1D is a top plan view: Structural characteristics of a trimeric soluble sTNF RII receptor-C-propeptide fusion. Given its 3-fold symmetry, a sTNF RII-Trimer fusion protein is trivalent in nature, thus can perfectly dock to its trimeric ligand TNF- $\alpha$ . C-propeptide of collagen capable of self trimerization is labeled 5 with inter-chain disulfide bonds 3.

Fig. 2A and Fig. 2B illustrate the structures of pTRIMER plasmid vectors for creating secreted trimeric fusion proteins. Any soluble receptor- or biological active polypeptide-encoding cDNAs can be cloned into the unique Hind III or Bgl II sites to allow in-frame fusion at the C-termini to the  $\alpha$  (I) collagen containing C-propeptide sequence for trimerization. Fig. 2A: The pTRIMER(T0) construct contains part of the glycine-repeats (GXY) $_n$  upstream of the C-propeptide; Fig. 2B: whereas the pTRIMER(T2) contains only the C-propeptide domain with a mutated BMP-1 protease recognition site.

Fig. 3A and Fig. 3B illustrate the expression and secretion of disulfide bond-linked trimeric collagen fusion proteins.

Fig. 3A. Western blot analysis of the trimerization of human placental alkaline phosphatase (AP) when fused to the C-propeptides of  $\alpha$ (I) collagen. The expression vectors encoding either AP alone or AP-C-propeptide fusions in pTRIMER vectors were transiently transfected into HEK293T cells. Forty-eight hours later, the

conditioned media (20  $\mu$ L) of each transfected cells as indicated were boiled for 5 minutes in equal volume of 2X SDS sample buffer either with or without reducing agent (mercaptoethanol), separated on a 10% SDS-PAGE and analyzed by Western blot using a polyclonal antibody to AP (GenHunter Corporation). Note the secreted  
5 67 kDa AP alone does not form intermolecular disulfide bonds, whereas the secreted AP-T0 and AP-T2 fusions both are assembled efficiently into disulfide bond linked trimers.

**Fig. 3B.** Western blot analysis of the trimerization of soluble human TNF-RII when fused to the C-propeptides of  $\alpha$ (I) collagen. The expression vectors encoding either  
10 the AP—C-propeptide fusion (T2) (as a negative control for antibody specificity), or human soluble TNF-RII-C-propeptide fusions as indicated in pTRIMER vectors were transiently transfected into HEK293T cells. Forty-eight hours later, the conditioned media (20  $\mu$ L) of each non-transfected and transfected cells as indicated were boiled for 5 minutes in equal volume of 2X SDS sample buffer either with or without  
15 reducing agent (mercaptoethanol), separated on a 10% SDS-PAGE and analyzed by Western blot using a monoclonal antibody to human TNF-RII (clone 226, R & D Systems, Inc.). Note the monoclonal antibody can only recognize the secreted TNF-RII with disulfide bonds. Both the soluble TNF-RII-T0 and TNF-RII-T2 fusions are assembled efficiently into disulfide bond linked trimers.

20 **Fig. 4 and Fig. 5** illustrate the bioassays showing the potent neutralizing activity of the trimeric soluble human TNF-RII-C-propeptide fusion protein against human TNF- $\alpha$  mediated apoptosis.

**Fig. 4.** The TNF- $\alpha$  sensitive WEHI-13VAR cells (ATCC) were resuspended at 1 million cells/mL in RPMI medium containing 10% FBS. 100  $\mu$ L of the cell  
25 suspension was plated into each well in a 96-well microtiter plate. Actinomycin D

was added to each well at 500 ng/mL concentration followed by human TNF- $\alpha$  at 500 pg/ml (R & D Systems) in the presence or absence of trimeric soluble human TNF-RII-T2 as indicated. As a negative control, the trimeric AP-T2 was added in place of TNF-RII-T2. After 16 hours of incubation in a tissue culture incubator, the viability of cells was examined using either an inverted microscope at 20X magnification or cell viability indicator dye, Alamar Blue (BioSource, Inc.) added to 10% (v/v) to each well. The live cells are able to turn the dye color from blue to pink. Note that the trimeric soluble human TNF-RII-T2 exhibits a potent neutralizing activity against TNF- $\alpha$  protects the cells from TNF- $\alpha$  mediated apoptosis

10

**Fig. 5.** Quantitative analysis of the neutralizing activity of trimeric soluble human TNF-RII-T2 against human TNF- $\alpha$ . The experiment was carried out as Fig. 4A. Two hours after adding the Alamar Blue dye, the culture medium as indicted from each well was analyzed at OD575. The readings were normalized against wells with either no TNF- $\alpha$  (100% viability) added or with TNF- $\alpha$  without neutralizing agent (0% viability) added.

15

## BRIEF DESCRIPTION OF SEQUENCE LISTINGS

## SEQ ID NO: 1 (963 bases)

Nucleotide sequence encoding the C-propeptide human collagen  $\alpha(I)$  T0 construct.

- The cDNA construct was cloned into the pAPtag2 vector, replacing the AP coding  
5 region. The underlined sequences denote restriction enzyme sites used in constructing  
the corresponding pTRIMER vector. The bolded codons denote the start and the stop  
of the T0 coding region.

## SEQ ID NO: 2 (311 aa)

- The predicted C-propeptide T0 protein sequence of human Collagen  $\alpha(I)$ . The  
10 underlined sequence denotes the region of the "glycine repeats" upstream of the C-  
propeptide. The amino acid residues in red indicate the BMP-1 protease recognition  
site.

## SEQ ID NO: 3 (771 bases)

Nucleotide sequence encoding the C-propeptide of human collagen  $\alpha(I)$  T2 construct.

- 15 The cDNA construct was cloned into pAPtag2 vector, replacing the AP coding region.  
The underlined sequences denote restriction enzyme sites used in constructing the  
corresponding pTRIMER vector. The bolded codons denote the start and the stop of  
the T2 coding region.

## SEQ ID NO: 4 (247 aa)

- 20 The predicted C-propeptide T2 protein sequence of human Collagen  $\alpha(I)$ . The amino  
acid residue in red indicates the location of mutated BMP-1 protease recognition site.

## SEQ ID NO: 5 (2487 bases)

Nucleotide sequence encoding the human placental alkaline phosphatase (AP) fused  
to the T0 C-propeptide of human  $\alpha(I)$  collagen (AP-T0). The underlined sequences



indicate the restriction sites used for the fusion construct. The restriction site, which marks the fusion site shown in the middle of the sequence, is Bgl II.

SEQ ID NO: 6 (819 aa)

The predicted protein sequence of the AP-T0 fusion protein. The amino acid residues in blue indicate fusion sites between human placental alkaline phosphates (AP) and the  $\alpha$ (I) collagen T0 polypeptide. The bolded codons denote the start and the stop of the fusion protein. The underlined sequence denotes the region of the "glycine repeats" upstream of the C-propeptide of human  $\alpha$ (I) collagen. The amino acid residues in red indicate the BMP-1 protease recognition sequence.

10 SEQ ID NO: 7 (2294 bases)

Nucleotide sequence encoding the human placental alkaline phosphatase (AP) fused to the T2 C-propeptide human  $\alpha$ (I) collagen (AP-T2). The bolded codons denote the start and the stop of the fusion protein. The underlined sequences indicate the restriction sites used for the fusion construct. The restriction site, which marks the fusion site shown in the middle of the sequence, is Bgl II.

SEQ ID NO: 8 (755 aa)

The predicted protein sequence of the AP-T2 Fusion. The amino acid residues in blue indicate fusion sites between human placental alkaline phosphates (AP) and the  $\alpha$ (I) collagen T2 polypeptide. The amino acid residue in red indicates the location of the mutated BMP-1 protease recognition site.

20 SEQ ID NO: 9 (1734 bases)

Nucleotide sequence encoding the human soluble TNF-RII fused to the T0 C-propeptide of human  $\alpha$ (I) collagen (sTNF-RII-T0). The bolded codons denote the start and the stop of the fusion protein. The underlined sequences indicate the restriction

sites used for the fusion construct. The underlined sequence, which marks the fusion site shown in the middle of the sequence, is the BamH I/Bgl II ligated junction.

SEQ ID NO: 10 (566 aa)

The predicted protein sequence of the human soluble TNF-RII-T0 Fusion. The amino acid residues in blue indicate fusion sites between human soluble TNF-RII and  $\alpha$ (I) collagen T0 polypeptide. The underlined sequence denotes region of the "glycine repeats" upstream of the C-propeptide of human  $\alpha$ (I) collagen. The amino acid residues in red indicate the BMP-1 protease recognition site.

SEQ ID NO: 11 (1542 bases)

10 Nucleotide sequence encoding the human soluble TNF-RII fused to the T2 C-propeptide of human  $\alpha$ (I) collagen (sTNF-RII-T2). The bolded codons denote the start and the stop of the fusion protein. The underlined sequences indicate the restriction sites used for the fusion construct. The underlined sequence, which marks the fusion site shown in the middle of the sequence, is the BamH I/Bgl II ligated junction.

15 SEQ ID NO: 12 (502 aa)

The predicted protein sequence of the human soluble TNF-RII-T2 fusion protein. The amino acid residues in blue indicate fusion sites between human soluble TNF-RII and the  $\alpha$ (I) collagen T2 polypeptide. The amino acid residue in red indicates the location of mutated BMP-1 protease recognition site.

20 SEQ ID NO: 13 (2139 bases)

Nucleotide sequence encoding the human soluble CD4 fused to the T0 C-propeptide of human  $\alpha$ (I) collagen. The underlined sequences indicate the restriction sites used for the fusion construct. The underlined sequence, which marks the fusion site shown in the middle of the sequence, is the Bgl II site.

25 SEQ ID NO: 14 (699 aa)

The predicted Protein Sequence of the human soluble CD4-T0 Fusion. The amino acid residues in blue indicate fusion sites between human soluble CD4 and  $\alpha(I)$  collagen T0 polypeptide. The underlined sequence denotes the region of the "glycine repeats" upstream of the C-propeptide of human  $\alpha(I)$  collagen. The amino acid residues in red indicate the BMP-1 protease recognition site.

SEQ ID NO: 15 (1947 bases)

Nucleotide sequence encoding the human soluble CD4 fused to the T2 C-propeptide of human  $\alpha(I)$  collagen. The underlined sequences indicate the restriction sites used for the fusion construct. The underlined sequence, which marks the fusion site shown in the middle of the sequence, is the Bgl II site.

SEQ ID NO: 16 (635 aa)

The predicted Protein Sequence of the human soluble CD4-T2 Fusion. The amino acid residues in blue indicate fusion sites between human soluble CD4 and  $\alpha(I)$  collagen T2 polypeptide. The amino acid residue in red indicates the location of mutated BMP-1 protease recognition site.

#### DESCRIPTION OF THE INVENTION

Prior to setting forth the invention, it may be helpful to an understanding thereof to set forth definitions of certain terms to be used hereinafter.

**DNA Construct:** A DNA molecule, generally in the form of a plasmid or viral vector, either single- or double-stranded that has been modified through recombinant DNA technology to contain segments of DNA joined in a manner that as a whole would not otherwise exist in nature. DNA constructs contain the information necessary to direct the expression and/or secretion of the encoding protein of interest.

**Signal Peptide Sequence:** A stretch of amino acid sequence that acts to direct the secretion of a mature polypeptide or protein from a cell. Signal peptides are

characterized by a core of hydrophobic amino acids and are typically found at the amino termini of newly synthesized proteins to be secreted or anchored on the cell surface. The signal peptide is often cleaved from the mature protein during secretion. Such signal peptides contain processing sites that allow cleavage of the signal peptides from the mature proteins as it passes through the protein secretory pathway. A signal peptide sequence when linked to the amino terminus of another protein without a signal peptide can direct the secretion of the fused protein. Most of the secreted proteins, such as growth factors, peptide hormones, cytokines and membrane proteins, such as cell surface receptors, contain a signal peptide sequence when synthesized as a nascent protein.

Soluble receptor: The extracellular domain, in part or as a whole, of a cell surface receptor, which is capable of binding its ligand. Generally, it does not contain any internal stretch of hydrophobic amino acid sequence responsible for membrane anchoring.

C-propeptide of collagens: The C-terminal globular, and non-triple-helical domain of collagens, which is capable of self-assembly into trimers. In contrast to the triple helical region of collagens, the C-propeptide does not contain any glycine repeat sequence and is normally proteolytically removed from procollagen precursor upon procollagen secretion before collagen fibril formation.

Glycine repeats: The central linear triple helix forming region of collagen which contains hundreds of (Gly-X-Y)<sub>n</sub> repeats in amino acid sequence. These repeats are also rich in proline at X or/and Y positions. Upon the removal of N-and C-propeptides, the glycine-repeats containing collagen triple helices can assemble into higher order of insoluble collagen fibrils, which makes up the main component of the cell matrix.

cDNA: Stands for complementary DNA or DNA sequence complementary to messenger RNA. In general cDNA sequences do not contain any intron (non-protein coding) sequences.

Prior to this invention, nearly all therapeutic antibodies and soluble receptor-Fc fusion proteins, such as Enbrel, are dimeric in structure (Fig. 1). Although these molecules, compared to their monomeric counterparts, have been shown to bind their target antigens or ligands with increased avidity, it is predicted that they are still imperfect, due to structural constraints, to bind their targets that have a homotrimeric structure. Examples of such therapeutically important trimeric ligands include TNF family of cytokines and HIV coat protein gp120. Therefore, from a structural point of view, it will be desirable to be also able to generate trimeric soluble receptors or antibodies, which can perfectly dock to their target trimeric ligands or antigens (Fig. 1), and thereby completely block the ligand actions. Such trimeric soluble receptors or chimeric antibodies are expected to have the highest affinity to their targets and thus can be used more effectively and efficiently to treat diseases such as arthritis and AIDS.

This invention discloses ways for generating such secreted trimeric receptors and biological active proteins by fusing them to the C-propeptides of collagen, which are capable of self-assembly into trimers. The following are the advantages of this invention: (1) collagen is the most abundant protein secreted in the body of a mammal, constituting nearly 25% of the total protein in the body; (2) the major forms of collagen naturally occur as trimeric helices, with their globular C-propeptides responsible for the initiating of trimerization, which are subsequently proteolytically cleaved upon triple helix formation; (3) the cleaved soluble trimeric C-propeptide of collagen is found naturally at sub microgram/mL level in the blood of mammals; (4)

the linear triple helical region of collagen can be included as a linker or excluded as part of the fusion protein so the distance between a protein to be trimerized and the C-propeptide of collagen can be precisely adjusted to achieve an optimal biological activity; (5) the recognition site of BMP1 which cleaves the C-propeptide off the pro-collagen can be mutated or deleted to prevent the disruption of a trimeric fusion protein; (6) the C-proptide domain provides a universal affinity tag, which can be used for purification of any secreted fusion proteins created by this invention; (7) unlike the IgG1 Fc tag which is known to have other biological functions such as binding to its own cell surface receptors, the only known biological function of the C-propeptide of collagen is its ability to initiate trimerization of nascent pro-collagen chains and keep the newly made pro-collagen trimer soluble before assembly into insoluble cell matrix. These unique properties of the C-propeptide of collagen would predict that this unique trimerization tag is unlikely going to be toxic, or immunogenic, making it an ideal candidate for therapeutic applications.

To demonstrate the feasibility for making secreted trimeric fusion proteins, cDNA sequences encoding the entire C-propeptides of human  $\alpha 1$  (I) containing either some glycine-repeat triple helical region (T0 construct, sequence ID No. 1-2), or no glycine-repeat with a mutated BMP-1 recognition site (T2 construct, Sequence ID No. 3-4) were amplified by RT-PCR using EST clones purchased from the American Type Culture Collection (ATCC). The amplified cDNAs were each cloned as a Bgl II-XbaI fragment into the pAptag2 mammalian expression vector (GenHunter Corporation; Leder et al., 1996 and 1998), replacing the AP coding region (Fig. 2). The resulting vectors are called pTRIMER, versions T2 and T0, respectively. The vectors allow convenient in-frame fusion of any cDNA template encoding a soluble receptor or biologically active protein at the unique Hind III and Bgl II sites. Such

fusion proteins have the collagen trimerization tags located at the C termini, similar to native pro-collagens.

Example 1:

To demonstrate the feasibility of this invention, a cDNA encoding the human  
5 secreted placental alkaline phosphatase (AP), including its native signal peptide  
sequence, was cut out as a Hind III-Bgl II fragment from the pAPtag4 vector  
(GenHunter Corporation; Leder et al., 1996 and 1998) and cloned into the  
corresponding sites of the pTRIMER-T0 and pTRIMER-T2 vectors. The resulting  
AP-collagen fusion constructs (sequence ID No. 5-8) were expressed in HEK293T  
10 cells (GenHunter Corporation) after transfection. The successful secretion of the AP-  
collagen fusion proteins can be readily determined by AP activity assay using the  
conditioned media of the transfected cells. The AP activity reached about 1 unit/mL  
(or equivalent to about 1  $\mu$ g/mL of the fusion protein) 2 days following the  
transfection. To obtain HEK293T cells stably expressing the fusion proteins, stable  
15 clones were selected following co-transfection with a puromycin-resistant vector,  
pBabe-Puro (GenHunter Corporation). Clones expressing AP activity were expanded  
and saved for long-term production of the fusion proteins.

To determine if the AP-collagen fusion proteins are assembled into disulfide  
bond-linked trimers, conditioned media containing either AP alone or AP-T0 and AP-  
20 T2 fusions were boiled in SDS sample buffers containing either without (non-  
reducing) or with  $\beta$ -mercaptoethanol (reducing), separated by an SDS PAGE and  
analyzed by Western blot using an anti-AP polycloning antibody (GenHunter  
Corporation). AP alone without fusion exhibited as a 67 kDa band under both non-  
reducing and reducing conditions, consistent with the lack of any inter-molecular  
25 disulfide bonds as expected (Fig. 3A). In contrast, both AP-T0 and AP-T2 fusion

proteins secreted were shown to be three times as big (about 300 kDa) under the non-reducing condition as those under the reducing condition (90-100 kDa), indicating that both fusion proteins were assembled completely into homotrimers (Fig. 3A). This result essentially reduces the concept of this invention to practice.

5 Example 2:

To provide a proof that new and therapeutically beneficial biological functions can be endowed to a trimeric fusion protein, next a trimeric human soluble TNF-RII (p75) receptor using a corresponding EST clone purchased from the ATCC was constructed. As described in Example 1, the N-terminal region of human TNF-RII, including the entire ligand-binding region, but excluding the trans-membrane domain, was cloned in-frame, as a Bam H I fragment, into the Bgl II site of both pTRIMER-T0 and pTRIMER-T2 vectors (Sequence ID Nos. 9-12). The resulting fusion constructs were expressed in HEK293T cells following transfection. Stable clones were obtained by puromycin co-selection as described in Example 1. Western blot analysis under both non-reducing and reducing conditions was carried out to determine if the resulting soluble TNF-RII-collagen fusion proteins were indeed expressed, secreted and assembled into trimeric forms. As expected, the monoclonal antibody against human TNF-RII (clone 226 from R & D Systems, Inc.) clearly recognized the trimeric soluble TNF-fusion proteins expressed by both T0 and T2 fusion vectors as 220-240 kDa bands, which are about three times bigger than the corresponding monomeric fusion proteins (Fig. 3B). The TNF-RII antibody failed to detect monomeric fusion proteins under reducing conditions, consistent with the property specified by the antibody manufacturer. As a negative control for antibody specificity, neither the HEK293T cell alone, nor the cells expressing AP-T2 fusion protein expressed any TNF-RII (Fig. 3B).



To determine if the trimeric soluble TNF-RII receptors are potent inhibitors of its trimeric ligand TNF- $\alpha$ , TNF- $\alpha$  bioassay was carried out using a cytokine sensitive cell line WEHI-13VAR (ATCC) essentially as described previously (Mohler et al., 1993). The result shown in Fig. 4 clearly indicated that the trimeric soluble TNF-RII-  
5 C-propeptide fusion proteins are extremely potent in neutralizing the TNF- $\alpha$  mediated apoptosis of WEHI-13VAR cells in the presence of Actinomycin D (500 ng/mL) (Sigma). When human TNF- $\alpha$  (R & D Systems) was used at 0.5 ng/mL, the trimeric soluble TNF-RII-T2 (both from serum-free media or in purified form) had an apparent Ki-50 (50% inhibition) of about 2 ng/mL or  $8 \times 10^{-12}$  M (assuming the MW of 240  
10 kDa as homotrimer). This affinity to TNF- $\alpha$  is 4 orders of magnitude higher than that of the monomeric TNF-RII and at least 10-100 times higher than that of the dimeric soluble TNF-RII-Fc fusion, such as Enbrel (Mohler et al., 1993).

This crucial example proves that this invention can create trimeric fusion proteins with new biological properties that may have great therapeutic applications.  
15 Such soluble trimeric human TNF receptors may prove to be much more effective than the current dimeric soluble TNF receptor (e.g. Enbrel) on the market in treating autoimmune diseases such as RA. The dramatically increased potency of trimeric-TNF receptors could greatly reduce the amount of TNF blockers to be injected weekly for each patient, while improving the treatment and significantly lowering the cost for  
20 the patients. The improved potency of trimeric TNF receptors should also alleviate the current bottleneck in dimeric TNF receptor production, which currently can only meet the demands in treating about 100,000 patients in the United States.

#### Example 3.

The HIV virus, the cause of AIDS, infects and destructs primarily a special lineage of  
25 T lymphocytes in our body. These so called CD4+ T cells express a cell surface

protein dubbed CD4, which is the receptor of HIV. HIV recognizes the CD4+ cells with its viral coat protein gp120 that binds to CD4. Notably, the gp120 exists as a giant homotrimeric complex on the viral surface, whereas the CD4 is monomeric on the cell surface. The current model for HIV infection is that of a complete docking of HIV to CD4+ T cells, when all three subunits of gp120 trimers are each bound to CD4 is required for viral RNA entry into the cells. Obviously, one of the straightforward strategies for stopping HIV infection is to use soluble CD4 to blind the virus. Indeed, such approach using both monomeric soluble CD4 and CD4-Fc fusions has been shown quite effective in curbing HIV infections of laboratory isolates (Clapham et al., 1989; Daar et al., 1990). Unfortunately, these soluble CD4 were less effective in stopping the infection of HIV viral strains found in AIDS patients (Daar et al., 1990), possibly due to the amino acid sequence variations of the gp120, which lowers the affinity to monomeric and dimeric soluble CD4s.

To significantly increase the affinity of a soluble CD4 to any gp120 variants on HIV viruses, ideally a soluble CD4 should be in trimeric form so it can perfectly dock to its trimeric ligand, gp120 homotrimers. One of the major challenges for combating AIDS has been the high mutational rate of the viral genome, which leads to drug resistance. Therefore any drugs that directly target viral genes, such as HIV reverse transcriptase (e.g. AZT) and protease, are likely rendered ineffective as a result of viral mutations. In contrast, no matter how much it mutates, a HIV virus has to bind to a cellular CD4 receptor to initiate the infection. Thus, a high affinity soluble CD4 trimer should be immune to viral mutations because viral mutations in gp120 genes will render the virus unable to bind not only to a trimeric soluble CD4, but also CD4 on the cells.

To create such trimeric soluble CD4 HIV receptor analogs, a cDNA encoding the entire human soluble CD4, including its native signal peptide sequence, but excluding the transmembrane and the short cytoplasmic domains, was amplified using an EST clone purchased from the ATCC. The resulting cDNA was then cloned as a  
5 Hind III-Bgl II fragment into the corresponding sites of the pTRIMER-T0 and pTRIMER-T2 expression vectors. The resulting soluble CD4-collagen fusion constructs (sequence ID No. 13-16) were expressed in HEK293T cells (GenHunter Corporation) after transfection. To obtain HEK293T cells stably expressing the fusion proteins, stable clones were selected following co-transfection with a  
10 puromycin-resistant vector, pBabe-Puro (GenHunter Corporation). Clones expressing the fusion proteins were expanded and saved for long-term production of the fusion proteins.

To determine if the soluble human CD4-collagen fusion proteins are assembled into disulfide bond-linked trimers, conditioned media containing soluble CD4-T0 and  
15 CD4-T2 fusions were boiled in SDS sample buffers containing either without (non-reducing) or with  $\beta$ -mercaptoethanol (reducing), separated by a SDS PAGE and analyzed by Western blot using an monoclonal antibody to human CD4 (R & D Systems). Both soluble CD4-T0 and CD4-T2 fusion proteins secreted were shown to be three times as big (about 300 kDa) under the non-reducing condition as those under  
20 the reducing condition (90-100 kDa), indicating they were assembled essentially completely into homotrimers (data not shown). Now these trimeric soluble CD4 can be readily tested for gp120 binding and anti-HIV infection.

## CLAIMS

What is claimed is:

5

1. A method for generating a secreted trimeric fusion protein, comprising:

- (a) creating a DNA construct comprising a transcriptional promoter linked to a template encoding a signal peptide sequence followed by in-frame fusion to polypeptide to be trimerized, which in turn is joined in-frame to a polypeptide capable of self-trimerization which is heterologous from the first polypeptide to be trimerized;
- 10 (b) introducing said DNA construct into a eukaryotic cell; (c) growing said host cell in an appropriate growth medium under physiological conditions to allow the secretion of a trimerized fusion-protein encoded by said DNA sequence; (d) isolating said trimerized fusion protein from the culture medium of said host cell.

15

2. The method of claim 1 wherein the trimerized polypeptide fusion is a homotrimer.

3. The method of claim 1 wherein the trimerizing polypeptide moiety comprises the C terminal portion of collagen capable of self-assembly into a trimer selected from the
- 20 group consisting of pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI).

4. A method for generating a secreted trimeric fusion protein, comprising:

- (a) introducing into a eukaryotic host cell a first DNA construct comprising a
- 25 promoter which drives the transcription of an open reading frame consisting of a signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be trimerized, which in turn is joined in-frame to the C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I),

pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI),  
pro.alpha.2(XI) and pro.alpha.3(XI); (b) introducing into a eukaryotic host cell a  
second DNA construct comprising a promoter which drives the transcription of an  
open reading frame consisting of a second signal peptide sequence which is linked in-  
5 frame to a second non-collagen polypeptide to be trimerized, which in turn is joined  
in-frame to the second C-terminal portion of collagen capable of self-trimerization,  
selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III),  
pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI);  
(c) growing the host cell in an appropriate growth medium under physiological  
10 conditions to allow the secretion of a trimerized fusion protein encoded by said first  
and second DNA sequences; and (d) isolating the secreted trimeric fusion protein  
from the host cell.

5. A method for generating a secreted trimeric fusion protein, comprising:

15 (a) introducing into a eukaryotic host cell a first DNA construct comprising a  
promoter which drives the transcription of an open reading frame consisting of a  
signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be  
trimerized, which in turn is joined in-frame to the C-terminal portion of collagen  
capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I),  
20 pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI),  
pro.alpha.2(XI) and pro.alpha.3(XI); (b) introducing into a eukaryotic host cell a  
second DNA construct comprising a promoter which drives the transcription of an  
open reading frame consisting of a second signal peptide sequence which is linked in-  
frame to a second non-collagen polypeptide to be trimerized, which in turn is joined  
25 in-frame to a second C-terminal portion of collagen capable of self-trimerization,

- selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III),  
pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI);  
(c) introducing into a eukaryotic host cell a third DNA construct comprising a  
promoter which drives the transcription of an open reading frame consisting of a third  
5 signal peptide sequence which is linked in-frame to a third non-collagen polypeptide  
to be trimerized, which in turn is joined in-frame to a third C-terminal portion of  
collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I),  
pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI),  
pro.alpha.2(XI) and pro.alpha.3(XI); (d) growing the host cell in an appropriate  
10 growth medium under physiological conditions to allow the secretion of a trimerized  
fusion protein encoded by said first and second DNA sequences; and (e) isolating the  
secreted trimeric fusion protein from the host cell.
6. The methods of claims 1-5, wherein the signal peptide sequence and the non-  
15 collagen polypeptide to be trimerized are both from the same native secreted protein.
7. The methods of claims 1- 5, wherein the signal peptide sequence and the non-  
collagenous protein to be trimerized are selected from two different secreted proteins.
- 20 8. The methods of claim 1, 4 and 5, wherein the host eukaryotic cell is a fungal or insect cell.
9. The methods of claim 1, 4 and 5, wherein the host eukaryotic cell is a cultured  
mammalian cell line.

10. The methods of claims 1-5, wherein a C-terminal portion of collagen includes a "glycine-repeat" triple helical region of collagen linked to a C-propeptide.

11. The methods of claim 10, wherein a C-terminal portion of collagen is identified by  
5 Sequence ID Nos. 1-2.

12. The methods of claims 1-5, wherein the trimerizing C-terminal portion of collagen comprises only a C-propeptide without any glycine-repeat triple helical region of collagen.

10

13. The methods of claims 10 -12, wherein the trimerizing C-terminal portion of collagen comprises a mutated or deleted BMP-1 protease recognition sequence, thereby conferring the fusion proteins resistance to said protease degradation.

15 14. The methods of claims 12-13, wherein the trimerizing C-terminal portion of collagen is identified by sequence ID Nos. 3-4.

15. Compositions of fusion proteins generated by the methods of claims 1, 2, 3, 10, 11, 12, 13 and 14 are soluble trimeric human TNF- $\alpha$  receptor II (p75) identified by  
20 Sequence ID Nos. 9-12.

16. Compositions of fusion proteins generated by the methods of claims 1, 2, 3, 10, 11, 12, 13 and 14 are soluble trimeric human CD4 identified by Sequence ID Nos. 13-16.

25

17. Compositions of fusion proteins generated by the methods of claims 1, 2, 3, 10, 11, 12, 13 and 14 are soluble trimeric soluble human placental alkaline phosphatase identified by Sequence ID No. 5-8.
- 5 18. A trimerized polypeptide fusion comprising three polypeptide chains, each of said chains comprises a ligand-binding domain of a receptor joined to a C-propeptide of collagen, wherein trimerization of the polypeptide fusion results in enhancement of biological activity.
- 10 19. A method of blocking TNF- $\alpha$  biological activity using a trimerized soluble TNF- $\alpha$  receptor II generated by claim 15.



1/5

Side-View

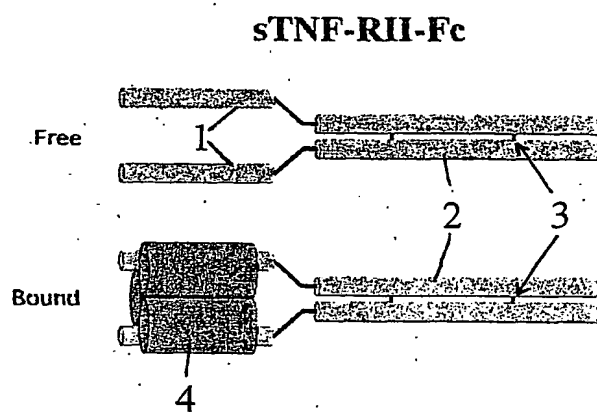


FIG. 1A

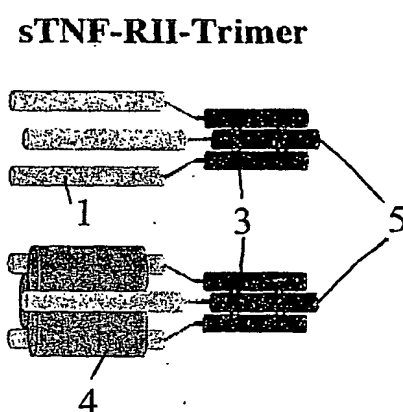


FIG. 1C

Top-View

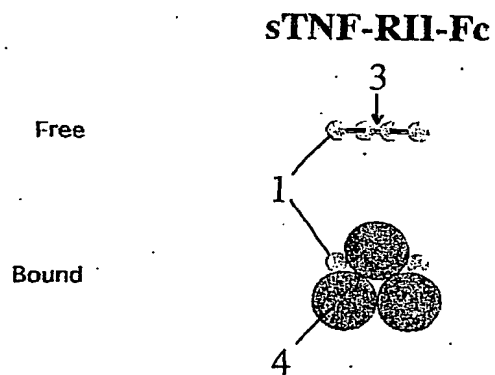


FIG. 1B

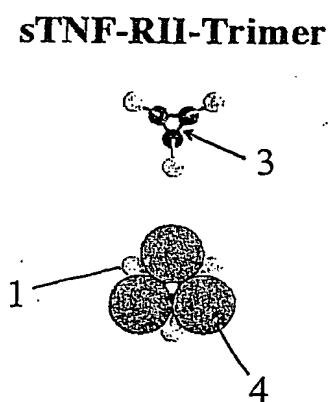


FIG. 1D

2/5

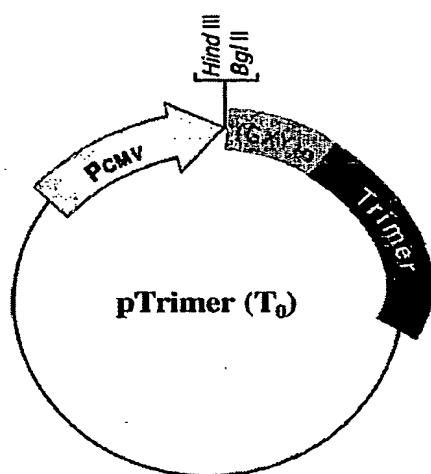


FIG. 2A

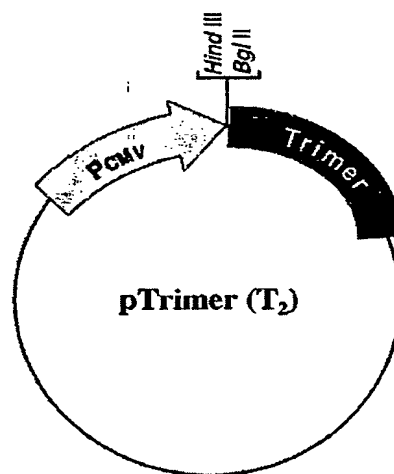


FIG. 2B

3/5

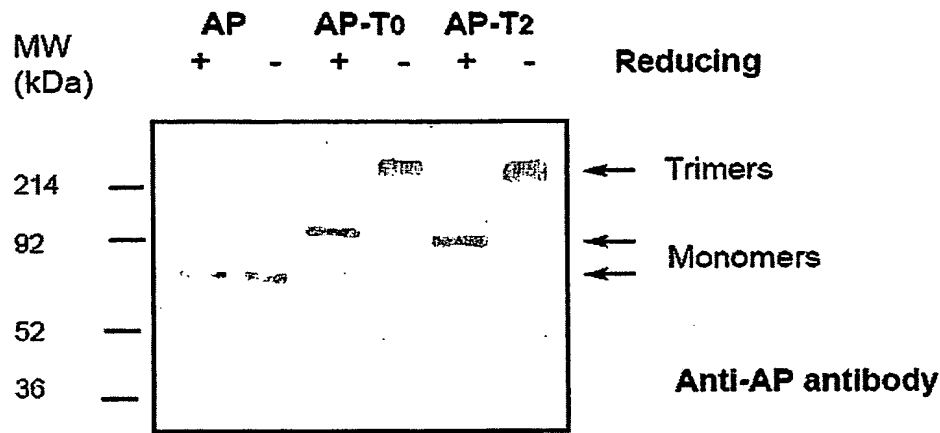


FIG. 3A

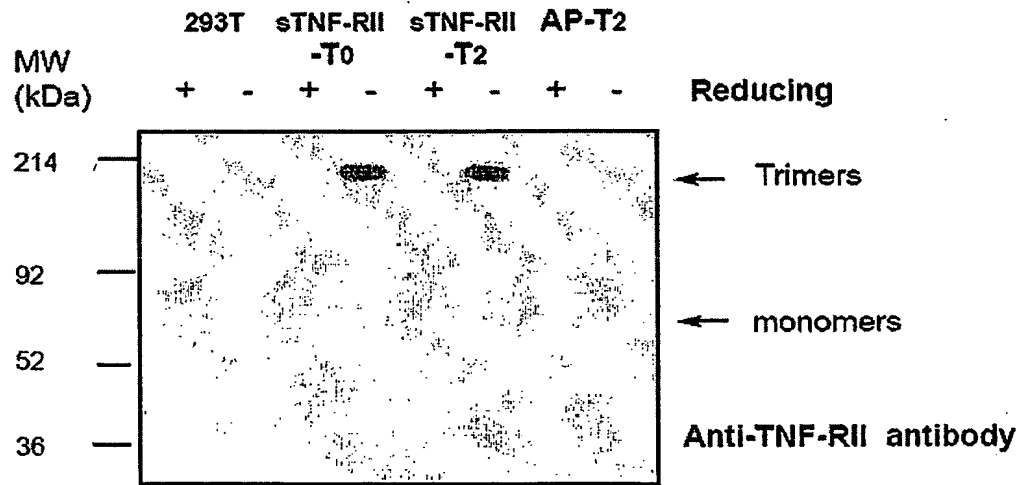


FIG. 3B

4/5

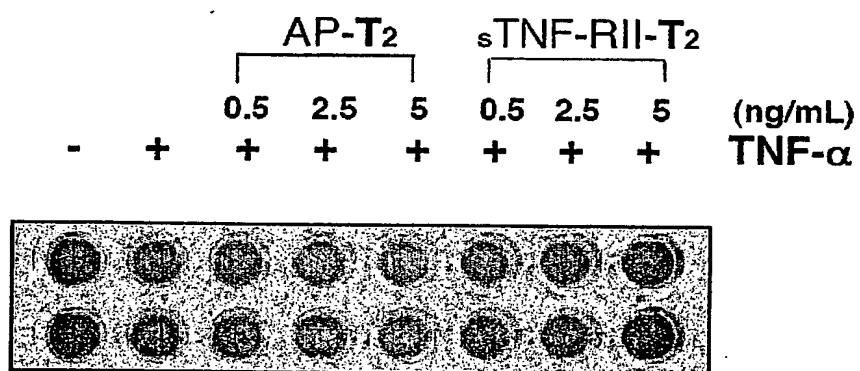


FIG. 4

5/5

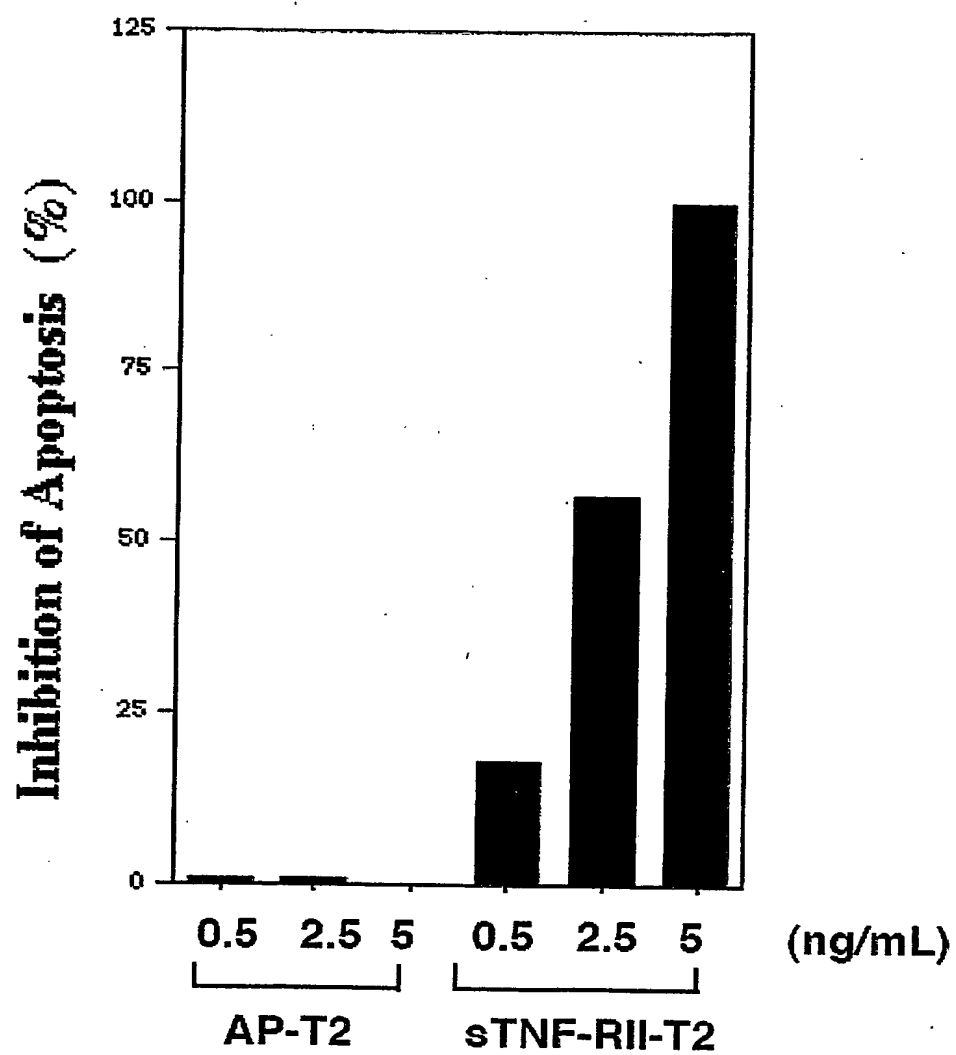


FIG. 5

## SEQUENCE LISTING

&lt;110&gt; GENHUNTER CORPORATION

<120> METHODS AND COMPOSITIONS FOR PRODUCING SECRETED  
TRIMERIC RECEPTOR ANALOGS AND BIOLOGICALLY ACTIVE FUSION  
PROTEINS

&lt;130&gt; 04-066-PL

&lt;150&gt; 10/677,877

&lt;160&gt; 16

&lt;210&gt; 1

&lt;211&gt; 963

&lt;212&gt; cDNA

&lt;213&gt; HOMO SAPIENS

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 12 947

&lt;400&gt; 1

Hind III

BglIII

AAGCTTACGTAAGATCTAACGGTCTCCCTGGCCCCATTGGGCCCCCTGGTCCT  
 CGCGGTGCGCACTGGTGATGCTGGTCCTGTTGGTCCCCCGGCCCTCCTGGACC  
 TCCTGGTCCCCCTGGTCCTCCCAGCGCTGGTTTCGACTTCAGCTTCCTGCCCC  
 AGCCACCTCAAGAGAAGGCTCACGATGGTGGCCGCTACTACCGGGCTGATGAT  
 GCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACCCTCAAGAGCCT  
 GAGCCAGCAGATCGAGAACATCCGGAGCCCAGAGGGAAGCCGCAAGAACCCCG  
 CCCGCACCTGCCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAG  
 TACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTG  
 CAACATGGAGACTGGTGAGACCTGCGTGTACCCCACTCAGCCCAGTGTGGCCO  
 AGAAGAAGTGGTACATCAGCAAGAACCCCAAGGACAAGAGGCATGTCTGGTTC  
 GGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGGCGGCCAGGGCTCCGA  
 CCCTGCCGATGTGGCCATCCAGCTGACCTTCCTGCGCCTGATGTCCACCGAGG  
 CCTCCCAGAACATCACCTACCACTGCAAGAACAGCGTGGCCTACATGGACCAG  
 CAGACTGGCAACCTCAAGAAGGCCCTGCTCCTCAAGGGCTCCAACGAGATCGA  
 GATCCGCGCCGAGGGCAACAGCCGCTTCACCTACAGCGTCACTGTGCGATGGCT  
 GCACGAGTCACACCGGAGCCTGGGGCAAGACAGTGATTGAATACAAAACCACC  
 AAGTCCTCCCGCCTGCCCATCATCGATGTGGCCCCCTTGGACGTTGGTGCCCC  
 AGACCAGGAATTCGGCTTCGACGTTGGCCCTGTCTGCTTCCTGTAAACTCCCT  
 CCATCTAGA

Xba I

&lt;210&gt; 2

&lt;211&gt; 311

&lt;212&gt; PROTEIN

&lt;213&gt; HOMO SAPIENS

&lt;400&gt; 2

1 RSNGLPPIG PPGPRGRTGD AGPVGPPGPP GPPGPPGPPS AGFDFSFLPQ PPQEK AHDGG 60  
 61 RYYRADANV VRDRDLEVD TLKSLSQIE NIRSPEGSRK NPARTCRDLK MCHSDWKSGE 120  
 121 YWIDPNQGCN LDAIKVFCNM ETGETCVYPT QPSVAQKNWY ISKNPKDKRH VWFGE SMTDG 180  
 181 FQFEYGGQGS DPADVAIQLT FLRLMSTEAS QNITYHCKNS VAYMDQQTGN LKKALLLKGS 240  
 241 NEIEIRAEGN SRFTYSVTVD GCTSHTGAWG KTVIEYKTTK SSRLPIIDVA PLDVGAPDQE 300  
 301 FGFDVGPVCF L

&lt;210&gt; 3

&lt;211&gt; 711

&lt;212&gt; cDNA

&lt;213&gt; HOMO SAPIENS

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 12 755

&lt;400&gt; 3

Hind III

BglIII

AAGCTTACGTAAGATCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACC  
 CTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCGGAGCCCAGAGGGAAGCCGCAAGAACCC  
 CGCCCGCACCTGCCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGA  
 TTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGT  
 GAGACCTGCGTGTACCCCACTCAGCCCAGTGTGGCCCAAGAAGTGGTACATCAGCAAGAA  
 CCCCAGGACAAGAGGCATGTCTGGTTCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGT  
 ATGGCGGCCAGGGCTCCGACCCTGCCGATGTGGCCATCCAGCTGACCTTCCTGCGCCTGATG  
 TCCACCGAGGCCTCCCAGAACATCACCTACCACTGCAAGAACAGCGTGGCCTACATGGACCA  
 GCAGACTGGCAACCTCAAGAAGGCCCTGCTCCTCAAGGGCTCCAACGAGATCGAGATCCGCG  
 CCGAGGGCAACAGCCGCTTCACCTACAGCGTCACTGTGATGGCTGCACGAGTCACACCGGA  
 GCCTGGGGCAAGACAGTGATTGAATACAAAACCACCAAGTCCTCCCGCCTGCCCATCATCGA  
 TGTGGCCCCCTTGGACGTTGGTGCCCCAGACCAGGAATTCGGCTTCGACGTTGGCCCTGTCT  
 GCTTCCTGTAAACTCCCTCCATCTAGA

Xba I

&lt;210&gt; 4

&lt;211&gt; 247

&lt;212&gt; PROTEIN

&lt;213&gt; HOMO SAPIENS

&lt;400&gt; 4

1 RSDANVVRDR DLEVDITLKS LSQQIENIRS PEGSRKNPAR TCRDLKMCHS DWKSGEYWID 60  
 61 PNQGCNLDAI KVFCNMETGE TCVYPTQPSV AQKNWYISK NPKDKRHVWFG ESMTDGFQFE 120  
 121 YGGQGSDPAD VAIQLTFLRL MSTEASQNT YHCKNSVAYM DQQTGNLKA LLLKGSNEIE 180  
 181 IRAEGNSRFT YSVTVDGCTS HTGAWGKTVI EYKTTKSSRL PIIDVAPLDV GAPDQEFQFD 240  
 241 VGPVCFL

&lt;210&gt; 5

&lt;211&gt; 2487

&lt;212&gt; cDNA

&lt;213&gt; HOMO SAPIENS

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 12 2471

&lt;400&gt; 5

Hind III

AAGCTTCCTGCATGCTGCTGCTGCTGCTGCTGCTGGGCCTGAGGCTACAGCTCTCCC  
 TGGGCATCATCCCAGTTGAGGAGGAGAACCCGGA CTCTGGAACCGCGAGGCAGCCG  
 AGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGCCGCCAAGAACCTCA  
 TCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTAA  
 AAGGGCAGAAGAAGGACAAACTGGGGCCTGAGATACCCCTGGCCATGGACCGCTTCC

CATATGTGGCTCTGTCCAAGACATACAATGTAGACAAACATGTGCCAGACAGTGGAG  
CCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCAGACCATTGGCTTGA  
GTGCAGCCGCCCCGCTTTAACCAGTGCAACACGACACGCGGCAACGAGGTTCATCTCCG  
TGATGAATCGGGCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCACACGAG  
TGCAGCACGCCTCGCCAGCCGGCACCTACGCCCACACGGTGAACCGCAACTGGTACT  
CGGACGCCGACGTGCCTGCCTCGGCCCCGCCAGGAGGGGTGCCAGGACATCGCTACGC  
AGCTCATCTCCAACATGGACATTGACGTGATCCTAGGTGGAGGCCGAAAGTACATGT  
TTCCCATGGGAACCCCAGACCCTGAGTACCCAGATGACTACAGCCAAGGTGGGACCA  
GGCTGGACGGGAAGAATCTGGTGCAGGAATGGCTGGCGAAGCGCCAGGGTGCCCGGT  
ATGTGTGGAACCGCACTGAGCTCATGCAGGCTTCCCTGGACCCGTCTGTGACCCATC  
TCATGGGTCTCTTTGAGCCTGGAGACATGAAATACGAGATCCACCGAGACTCCACAC  
TGGACCCCTCCCTGATGGAGATGACAGAGGCTGCCCTGCGCCTGCTGAGCAGGAACC  
CCCGCGGCTTCTTCCCTCTTCGTGGAGGGTGGTTCGCATCGACCATGGTCATCATGAAA  
GCAGGGCTTACCGGGCACTGACTGAGACGATCATGTTTCGACGACGCCATTGAGAGGG  
CGGGCCAGCTCACCAGCGAGGAGGACACGCTGAGCCTCGTCACTGCCGACCACTCCC  
ACGTCTTCTCCTTCGGAGGCTACCCCCCTGCGAGGGAGCTCCATCTTCGGGCTGGCCC  
CTGGCAAGGCCCGGGACAGGAAGGCCTACACGGTCCTCCTATACGGAAACGGTCCAG  
GCTATGTGCTCAAGGACGGCGCCCCGGCCGATGTTACCGAGAGCGAGAGCGGGAGCC  
CCGAGTATCGGCAGCAGTCAGCAGTGCCCCCTGGACGAAGAGACCCACGCAGGCGAGG  
ACGTGGCGGTGTTTCGCGCGCGGCCCCGAGGCGCACCTGGTTTCACGGCGTGCAGGAGC  
AGACCTTCATAGCGCACGTTCATGGCCTTCGCGCCTGCCTGGAGCCCTACACCGCCT  
GCGACCTGGCGCCCCCGCCGGCACCAACCGACGCCGCGCACCCGGGTTCCGGAAGAT  
CTAACGGTCTCCCTGGCCCCATTGGGCCCCCTGGTCCTCGCGGTGCGACTGGTGATG  
CTGGTCCTGTTGGTCCCCCGGCCCTCCTGGACCTCCTGGTCCCCCTGGTCCTCCCA  
GCGCTGGTTTTGACTTCAGCTTCCTGCCCCAGCCACCTCAAGAGAAGGCTCACGATG  
GTGGCCGCTACTACCGGGCTGATGATGCCAATGTGGTTTCGTGACCGTGACCTCGAGG  
TGGACACCACCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCGGAGCCCAGAGG  
GAAGCCGCAAGAACCCCGCCCCGACCTGCCGTGACCTCAAGATGTGCCACTCTGACT  
GGAAGAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCA  
AAGTCTTCTGCAACATGGAGACTGGTGAGACCTGCGTGTACCCCACTCAGCCCAGTG  
TGGCCCAGAAGAACTGGTACATCAGCAAGAACCCCAAGGACAAGAGGCATGTCTGGT  
TCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGGCGGCCAGGGCTCCGACC  
CTGCCGATGTGGCCATCCAGCTGACCTTCCTGCGCCTGATGTCCACCGAGGCCTCCC  
AGAACATCACCTACCACTGCAAGAACAGCGTGGCCTACATGGACCAGCAGACTGGCA  
ACCTCAAGAAGGCCCTGCTCCTCAAGGGCTCCAACGAGATCGAGATCCGCGCCGAGG  
GCAACAGCCGCTTCACCTACAGCGTCACTGTGATGGCTGCACGAGTCACACCGGAG  
CCTGGGGCAAGACAGTGATTGAATACAAAACCAAGTCCTCCCGCCTGCCCATCA  
TCGATGTGGCCCCCTTGGACGTTGGTGCCCCAGACCAGGAATTTCGGCTTCGACGTTG  
GCCCTGTCTGCTTCCTGTAAACTCCCTCCATCTAGA

Xba I

&lt;210&gt; 6

&lt;211&gt; 819

&lt;212&gt; PROTEIN

&lt;213&gt; HOMO SAPIENS

&lt;400&gt; 6

1 MLLLLLLLGL RLQLSLGIIP VEEENPDFWN REAAEALGAA KKLQPAQTAA KNLIFLGDG 60  
61 MGVSSTVTAAR ILKGQKKDKL GPEIPLAMDR FPYVALSKTY NVDKHPVDSG ATATAYLCGV 120



121 KGNFQTIGLS AAARFNQCNT TRGNEVISVM NRAKKAGKSV GVVTTTRVQH ASPAGTYAHT 180  
 181 VNRNWYSDAD VPASARQEGC QDIATQLISN MDIDVILGGG RKYMFPMGTP DPEYPDDYSQ 240  
 241 GGTRLDGKNL VQEWLAKRQG ARYVWNRTEL MQASLDPSVT HLMGLFEPGD MKYEIHRDST 300  
 301 LDPSLMEMTE AALRLLSRNP RGFFLFVEGG RIDHGHESR AYRALTETIM FDDAIERAGQ 360  
 361 LTSEEDTSL VTADHSHVFS FGGYPLRGSS IFGLAPGKAR DRKAYTVLLY GNGPGYVLKD 420  
 421 GARPDVTESE SGSPEYRQQS AVPLDEETHA GEDVAVFARG PQAHLVHGVQ EQTFIAHVMA 480  
 481 FAACLEPYTA CDLAPPAGTT DAAHPGSGRS NGLPGPIGPP GPRGRTGDAG PVGPPGPPGP 540  
 541 PGPPGPPSAG FDFSFLPQP QEKAHDGGY YRADDANVVR DRDLEVDTTL KSLSQIQIENI 600  
 601 RSPEGSRKNP ARTCRDLKMC HSDWKSGEYW IDPNQGCNLD AIKVFCNMET GETCVYPTQP 660  
 661 SVAQKNWYIS KNPDKRHVW FGESMTDGFQ FEYGGQSDP ADVAIQLTFL RLMSTEASQN 720  
 721 ITYHCKNSVA YMDQQTGNLK KALLKGSNE IEIRAEGNSR FTYSVTVDGC TSHTGAWGKT 780  
 781 VIEYKTTKSS RLPIDVAPL DVGAPDQEFQ FDVGPVCFL

&lt;210&gt; 7

&lt;211&gt; 2294

&lt;212&gt; cDNA

&lt;213&gt; HOMO SAPIENS

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 12 2278

&lt;400&gt; 7

Hind III

AAGCTTCCTGC**ATG**CTGCTGCTGCTGCTGCTGCTGCTGGGCCTGAGGCTACAGCTCTCCC  
 TGGGCATCATCCCAGTTGAGGAGGAGAACCCGGACTTCTGGAACCGCGAGGCAGCCG  
 AGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGCCGCCAAGAACCTCA  
 TCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTAA  
 AAGGGCAGAAGAAGGACAACTGGGGCCTGAGATAACCCCTGGCCATGGACCGCTTCC  
 CATATGTGGCTCTGTCCAAGACATACAATGTAGACAAACATGTGCCAGACAGTGGAG  
 CCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCAGACCATTGGCTTGA  
 GTGCAGCCGCCCCGCTTTAACCAGTGCAACACGACACGCGGCAACGAGGTCATCTCCG  
 TGATGAATCGGGCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCACACGAG  
 TGCAGCACGCCTCGCCAGCCGGCACCTACGCCCACACGGTGAACCGCAACTGGTACT  
 CGGACGCCGACGTGCCTGCCTCGGCCCCGCCAGGAGGGGTGCCAGGACATCGCTACGC  
 AGCTCATCTCCAACATGGACATTGACGTGATCCTAGGTGGAGGCCGAAAGTACATGT  
 TTCCCATGGGAACCCAGACCCTGAGTACCCAGATGACTACAGCCAAGGTGGGACCA  
 GGCTGGACGGGAAGAATCTGGTGCAGGAATGGCTGGCGAAGCGCCAGGGTGGCCGGT  
 ATGTGTGGAACCGCACTGAGCTCATGCAGGCTTCCCTGGACCCGTCTGTGACCCATC  
 TCATGGGTCTCTTTGAGCCTGGAGACATGAAATACGAGATCCACCGAGACTCCACAC  
 TGGACCCCTCCCTGATGGAGATGACAGAGGCTGCCCTGCGCCTGCTGAGCAGGAACC  
 CCCGCGGCTTCTTCCTCTTCGTGGAGGGTGGTGCATCGACCATGGTCATCATGAAA  
 GCAGGGCTTACCGGGCACTGACTGAGACGATCATGTTTCGACGACGCCATTGAGAGGG  
 CGGGCCAGCTCACCAGCGAGGAGGACACGCTGAGCCTCGTCACTGCCGACCACTCCC  
 ACGTCTTCTCCTTCGGAGGCTACCCCTGCGAGGGAGCTCCATCTTCGGGCTGGCCC  
 CTGGCAAGGCCCGGGACAGGAAGGCCTACACGGTCCTCCTATACGGAACGGTCCAG  
 GCTATGTGCTCAAGGACGGCGCCCGGCCGGATGTTACCGAGAGCGAGAGCGGGAGCC  
 CCGAGTATCGGCAGCAGTCAGCAGTGCCCTGGACGAAGAGACCCACGCAGGCGAGG  
 ACGTGGCGGTGTTTCGCGCGCGGCCCGCAGGCGCACCTGGTTTCACGGCGTGACGAGC  
 AGACCTTCATAGCGCACGTCATGGCCTTCGCCGCCTGCCTGGAGCCCTACACCGCCT  
 GCGACCTGGCGCCCCCGCCGGCACCAACCGACGCCGCGCACCCGGGTTCGGGAGATC

TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACCCTCAAGAGCCT  
 GAGCCAGCAGATCGAGAACATCCGGAGCCCAGAGGGAAGCCGCAAGAACCCCGCCCG  
 CACCTGCCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGAT  
 TGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGAC  
 TGGTGAGACCTGCGTGTACCCCACTCAGCCCAGTGTGGCCCAGAAGAACTGGTACAT  
 CAGCAAGAACCCCAAGGACAAGAGGCATGTCTGGTTCGGCGAGAGCATGACCGATGG  
 ATTCCAGTTCGAGTATGGCGGCCAGGGCTCCGACCCTGCCGATGTGGCCATCCAGCT  
 GACCTTCCTGCGCCTGATGTCCACCGAGGCCTCCCAGAACATCACCTACCACTGCAA  
 GAACAGCGTGGCCTACATGGACCAGCAGACTGGCAACCTCAAGAAGGCCCTGCTCCT  
 CAAGGGCTCCAACGAGATCGAGATCCGCGCCGAGGGCAACAGCCGCTTCACCTACAG  
 CGTCACTGTGCGATGGCTGCACGAGTCACACCGGAGCCTGGGGCAAGACAGTGATTGA  
 ATACAAAACCACCAAGTCTTCCCGCCTGCCCATCATCGATGTGGCCCCCTTGACGCT  
 TGGTGCCCCAGACCAGGAATTCGGCTTCGACGTTGGCCCTGTCTGCTTCCTGTAAAC  
 TCCCTCCATCTAGA

Xba I

&lt;210&gt; 8.

&lt;211&gt; 755

&lt;212&gt; PROTEIN

&lt;213&gt; HOMO SAPIENS

&lt;400&gt; 8

1 MLLLLLLLGL RLQLSLGHP VEEENPDFWN REAAEALGAA KKLOPAQTAA KNLHFLGDG 60  
 61 MGVSIVTAAR ILKGQKKDKL GPEIPLAMDR FPYVALSKTY NVDKHVPDSG ATATAYLCGV 120  
 121 KGNFQTIGLS AAARFNQCNTRGNEVISVM NRAKKAGKSV GVVTTTRVQH ASPAGTYAHT 180  
 181 VNRNWYSDAD VPASARQEGC QDIATQLISN MDIDVILGGG RKYMFPMGTP DPEYPDDYSQ 240  
 241 GGTRLDGKNL VQEWLAKRQG ARYVWNRTEL MQASLDPSVT HLMGLFEPGD MKYEIHRDST 300  
 301 LDPPLMEMTE AALRLLSRNP RGFFLFVEGG RIDHGHESR AYRALTETIM FDDAIERAGQ 360  
 361 LTSEEDTLSL VTADHSHVFS FGGYPLRGSS IFGLAPGKAR DRKAYTVLLY GNGPGYVLKD 420  
 421 GARPDVTESE SGSPEYRQQS AVPLDEETHA GEDVAVFARG PQAHLVHGVQ EQTFIAHVMA 480  
 481 FAACLEPYTA CDLAPPAGTT DAAHPGSGRS DANVVRDRDL EVDTLKLSLS QQIENIRSPE 540  
 541 GSRKNPARTC RDLKMCHSDW KSGEYWIDPN QGCNLDAIKV FCNMETGETC VYPTQPSVAQ 600  
 601 KNWYISKNP DKRHVWFGES MTDGFQFEYG GQGSDDADVA IQLTFLRLMS TEASQNITYH 660  
 661 CKNSVAYMDQ QTGNLKKALL LKGSNEIEIR AEGNSRFTYS VTDGCTSHT GAWGKTVIEY 720  
 721 KTTKSSRLPI IDVAPLDVGA PDQEFQFDVG PVCFL

&lt;210&gt; 9

&lt;211&gt; 1734

&lt;212&gt; cDNA

&lt;213&gt; HOMO SAPIENS

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 18 1718

&lt;400&gt; 9

Bam HI

GGATCCCGCCCGCACCCATGGCGCCCGCTCGCCGTCTGGGCGCGCTGGCCGTGCGACTGGAG  
 CTCTGGGCTGCGGCGCACGCCTTGCCCGCCAGGTGGCATTACACCCCTACGCCCGGAGCC  
 CGGGAGCACATGCCGGCTCAGAGAATACTATGACCAGACAGCTCAGATGTGCTGCAGCAAAT  
 GCTCGCCGGGCCAACATGCAAAAGTCTTCTGTACCAAGACCTCGGACACCGTGTGTGACTCC  
 TGTGAGGACAGCACATACCCAGCTCTGGAAGTGGGTTCCCGAGTGCTTGAGCTGTGGCTC  
 CCGCTGTAGCTCTGACCAGGTGAAACTCAAGCCTGCACTCGGGAACAGAACCAGCATCTGCA

CCTGCAGGCCCCGGCTGGTACTGCGCGCTGAGCAAGCAGGAGGGGTGCCGGCTGTGCGCGCCG  
 CTGCGCAAGTGCCGCCCCGGGCTTCGGCGTGGCCAGACCAGGAAGTAAACATCAGACGTGGT  
 GTGCAAGCCCTGTGCCCCGGGGACGTTCTCCAACACGACTTCATCCACGGATATTTGCAGGC  
 CCCACCAGATCTGTAACGTGGTGGCCATCCCTGGGAATGCAAGCATGGATGCAGTCTGCACG  
 TCCACGTCCCCACCCGGAGTATGGCCCCAGGGGCAGTACACTTACCCAGCCAGTGTCCAC  
 ACGATCCCAACACACGCAGCCAACCTCCAGAACCCAGCACTGCTCCAAGCACCTCCTTCCTGC  
 TCCCAATGGGCCCCAGCCCCCAGCTGAAGGGAGCACTGGATCTAACGGTCTCCCTGGCCC  
 CATTGGGGCCCCCTGGTCCTCGCGGTGCGACTGGTGGTGGTCTTGGTCCCCC  
 CGGCCCTCCTGGACCTCCTGGTCCCCCTGGTCTCCAGCGCTGGTTTCGACTTCAG  
 CTTCCTGCCCCAGCCACCTCAAGAGAAGGCTCACGATGGTGGCCGCTACTACCGGGC  
 TGATGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACCTCAAGAG  
 CCTGAGCCAGCAGATCGAGAACATCCGGAGCCCAGAGGGAAGCCGCAAGAACCCCGC  
 CCGCACCTGCCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTG  
 GATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGA  
 GACTGGTGGAGACCTGCGTGTACCCCACTCAGCCCAGTGTGGCCAGAGAAGTGGTA  
 CATCAGCAAGAACCCCAAGGACAAGAGGCATGTCTGGTTCGGCGAGAGCATGACCGA  
 TGGATTCCAGTTCGAGTATGGCGGCCAGGGCTCCGACCCTGCCGATGTGGCCATCCA  
 GCTGACCTTCCTGCGCCTGATGTCCACCGAGGCCTCCCAGAACATCACCTACCACTG  
 CAAGAACAGCGTGGCCTACATGGACCAGCAGACTGGCAACCTCAAGAAGGCCCTGCT  
 CCTCAAGGGCTCCAACGAGATCGAGATCCGCGCCGAGGGCAACAGCCGCTTCACCTA  
 CAGCGTCACTGTGATGGCTGCACGAGTCACACCGGAGCCTGGGGCAAGACAGTGAT  
 TGAATACAAAACCAAGTCTTCCCGCCTGCCCATCATCGATGTGGCCCCCTTGGA  
 CGTTGGTGGCCCAGACCAGGAATTCCGGCTTCGACGTTGGCCCTGTCTGCTTCCTGTA  
 AACTCCCTCCATCTAGA

Xba I

&lt;210&gt; 10

&lt;211&gt; 566

&lt;212&gt; PROTEIN

&lt;213&gt; HOMO SAPIENS

&lt;400&gt; 10

1 MAPVAVWAAL AVGLELWAAA HALPAQVAFT PYAPEPGSTC RLREYYDQTA QMCCSKCSPG 60  
 61 QHAKVFCTKT SDTVCDSCED STYTQLWNWV PECLSCGSRG SSDQVETQAC TREQNRICTC 120  
 121 RPGWYCALSK QEGCRLCAPL RKCRLPGFGVA RPTGTSVTV CKPCAPGTFS NTTSSDICT 180  
 181 PHQICNVVAI PGNASMDAVC TSTSPTRMA PGAVHLPQPV STRSQHTQPT PEPSTAPSTS 240  
 241 FLLPMGPSPP AEGSTGSNGL PGPIGPPGPR GRTGDAGPVG PPGPPGPPGP PGPPSAGFDF 300  
 301 SFLPQPPQEK AHDGGRYRYA DDANVVRDRD LEVDITLKS L SQQIENIRSP EGSRKNPART 360  
 361 CRDLKMCHSD WKSGEYWIDP NQGCNLDAIK VFCNMETGET CVYPTQPSVA QKNWYISKNP 420  
 421 KDKRHVWFGE SMTDGFQFEY GGQGSPPADV AIQLTFLRLM STEASQNTY HCKNSVAYMD 480  
 481 QQTGNLKKAL LLKGSNEIEI RAEGNSRFTY SVTVDGCTSH TGAWGKTVIE YKTTKSSRLP 540  
 541 IIDVAPLDVG APDQEFQFDV GPVCF

<210> 11  
 <211> 1542  
 <212> cDNA  
 <213> HOMO SAPIENS  
 <220>  
 <221> CDS  
 <222> 18 1526  
 <400> 11

Bam HI

GGATCCCGCCCGCACCC**ATGG**CGCCCGTCGCCGTCTGGGCGCGCTGGCCGTCGGACTGGAG  
 CTCTGGGCTGCGGCGCACGCCTTGCCCGCCAGGTGGCATTACACCCTACGCCCCGGAGCC  
 CGGGAGCACATGCCGGCTCAGAGAATACTATGACCAGACAGCTCAGATGTGCTGCAGCAAAT  
 GCTCGCCGGGCCAACATGCAAAAGTCTTCTGTACCAAGACCTCGGACACCGTGTGTGACTCC  
 TGTGAGGACAGCACATACACCCAGCTCTGGAAGTGGGTTCCCGAGTGCTTGAGCTGTGGCTC  
 CCGCTGTAGCTCTGACCAGGTGGAACTCAAGCCTGCACTCGGGAACAGAACC GCATCTGCA  
 CCTGCAGGCCCGGCTGGTACTGCGCGCTGAGCAAGCAGGAGGGGTGCCGGCTGTGCGCGCCG  
 CTGCGCAAGTGCCGCCCGGGCTTCGGCGTGGCCAGACCAGGAAGTGAACATCAGACGTGGT  
 GTGCAAGCCCTGTGCCCCGGGGACGTTCTCCAACACGACTTCATCCACGGATATTTGCAGGC  
 CCCACCAGATCTGTAACGTGGTGGCCATCCCTGGGAATGCAAGCATGGATGCAGTCTGCACG  
 TCCACGTCCCCACCCGGAGTATGGCCCCAGGGGAGTACACTTACCCCAGCCAGTGTCCAC  
 ACGATCCCAACACACGCAGCCAACTCCAGAACCCAGCACTGCTCCAAGCACCTCCTTCCTGC  
 TCCCAATGGGCCCCAGCCCCCAGCTGAAGGGAGCACTGGATCTGATGCCAATGTGGTTTCG  
 TGACCGTGACCTCGAGGTGGACACCACCTCAAGAGCCTGAGCCAGCAGATCGAGAA  
 CATCCGGAGCCCAGAGGGAAGCCGCAAGAACCCCGCCCGCACCTGCCGTGACCTCAA  
 GATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTG  
 CAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTGCGTGTA  
 CCCCACTCAGCCAGTGTGGCCAGAGAAGTGGTACATCAGCAAGAACCCCAAGGA  
 CAAGAGGCATGTCTGGTTTCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG  
 CGGCCAGGGCTCCGACCCTGCCGATGTGGCCATCCAGCTGACCTTCCTGCGCCTGAT  
 GTCCACCGAGGCCTCCCAGAACATCACCTACCACTGCAAGAACAGCGTGGCCTACAT  
 GGACCAGCAGACTGGCAACCTCAAGAAGGCCCTGCTCCTCAAGGGCTCCAACGAGAT  
 CGAGATCCGCGCCGAGGGCAACAGCCGCTTCACCTACAGCGTCACTGTGATGGCTG  
 CACGAGTCACACCGAGCCTGGGGCAAGACAGTGATTGAATACAAAACCAAGTC  
 CTCCCGCCTGCCCATCATCGATGTGGCCCCCTTGGACGTTGGTGCCCCAGACCAGGA  
 ATTTCGGCTTCGACGTTGGCCCTGTCTGCTTCCTG**TAA**ACTCCCTCCATCTAGA

Xba I

<210> 12  
 <211> 502  
 <212> PROTEIN  
 <213> HOMO SAPIENS  
 <400> 12

1 MAPVAVWAAL AVGLELWAAA HALPAQVAFT PYAPEPGSTC RLREYYDQTA QMCCSKCSPG 60  
 61 QHAKVFCTKT SDTVCDSCED STYTQLWNWV PECLSCGSRG SSDQVETQAC TREQNRITC 120  
 121 RPYWYCALSK QEGCRLCAPL RKCPRPGFVA RPTETSDVV CKPCAPGTFN NTTSSDIDR 180  
 181 PHQICNVVAI PGNASMDAVC TSTSPTRSMG PGAVHLPQPV STRSQHTQPT PEPSTAPSTS 240  
 241 FLLPMGPSPP AEGSTGSDAN VVRDRDLEVD TTLKSLSQQI ENIRSPGSR KNPARTCRDL 300  
 301 KMCHSDWKSG EYWIDPNQGC NLDAIKVFCN METGETCVYP TQPSVAQKNW YISKNPDKDR 360

361 HVWFGESMTD GFQFEYGGQG SDPADVAIQL TFLRLMSTEA SQNTYHCKN SVAYMDQQTG 420  
421 NLKKALLLK G SNEIEIRAEG NSRFTYSVTV DGCTSHTGAW GKTVIEYKTT KSSRLPIIDV 480  
481 APLDVGAPDQ EFGFDVGPVC FL

<210> 13

<211> 2139

<212> cDNA

<213> HOMO SAPIENS

<220>

<221> CDS

<222> 24 2123

<400> 13

Hind III

AAGCTTCCCTCGGCAAGGCCACAATGAACCGGGGAGTCCCTTTTAGGCACTTGCTTC  
TGGTGCTGCAACTGGCGCTCCTCCCAGCAGCCACTCAGGGAAAGAAAGTGGTGCTGG  
GCAAAAAGGGGATACAGTGGAAGTACCTGTACAGCTTCCCAGAAGAAGAGCATAC  
AATTCCTACTGGAAAACTCCAACCAGATAAAGATTCTGGGAAATCAGGGCTCCTTCT  
TAACTAAAGGTCCATCCAAGCTGAATGATCGCGCTGACTCAAGAAGAAGCCTTTGGG  
ACCAAGGAACTTTCCCCTGATCATCAAGAATCTTAAGATAGAAGACTCAGATACTT  
ACATCTGTGAAGTGGAGGACCAGAAGGAGGAGGTGCAATTGCTAGTGTTCGGATTGA  
CTGCCAACTCTGACACCCACCTGCTTCAGGGGCAGAGCCTGACCCTGACCTTGGAGA  
GCCCCCTGGTAGTAGCCCCCTCAGTGCAATGTAGGAGTCCAAGGGGTAAAAACATAC  
AGGGGGGGAAGACCCCTCTCCGTGTCTCAGCTGGAGCTCCAGGATAGTGGCACCTGGA  
CATGCACTGTCTTGCAAGAACAGAAGAGGTGGAGTTCAAAATAGACATCGTGGTGC  
TAGCTTTCCAGAAGGCCTCCAGCATAGTCTATAAGAAAGAGGGGGAACAGGTGGAGT  
TCTCCTTCCCCTCGCCTTTACAGTTGAAAAGCTGACGGGCAGTGGCGAGCTGTGGT  
GGCAGGCGGAGAGGGCTTCCTCCTCCAAGTCTTGATCACCTTTGACCTGAAGAACA  
AGGAAGTGTCTGTAAAACGGGTTACCCAGGACCCTAAGCTCCAGATGGGCAAGAAGC  
TCCCGCTCCACCTCACCTGCCCCAGGCCTTGCCCTCAGTATGCTGGCTCTGGAAACC  
TCACCTGGCCCTTGAAGCGAAAACAGGAAAGTTGCATCAGGAAGTGAACCTGGTGG  
TGATGAGAGCCACTCAGCTCCAGAAAAATTTGACCTGTGAGGTGTGGGGACCCACCT  
CCCCTAAGCTGATGCTGAGCTTGAACTGGAGAACAAAGGAGGCAAAGGTCTCGAAGC  
GGGAGAAGGCGGTGTGGGTGCTGAACCCTGAGGCGGGGATGTGGCAGTGTCTGCTGA  
GTGACTCGGGACAGGTCCTGCTGGAATCCAACATCAAGGTTCTGCCCAGATCTAACG  
GTCTCCCTGGCCCCATTGGGCCCCCTGGTCCTCGCGGTGCGCACTGGTGATGCTGGTC  
CTGTTGGTCCCCCGGCCCTCCTGGACCTCCTGGTCCCCCTGGTCCTCCCAGCGCTG  
GTTTCGACTTCAGCTTCCTGCCCCAGCCACCTCAAGAGAAGGCTCACGATGGTGGCC  
GCTACTACCGGGCTGATGATGCCAATGTGGTTCTGTGACCGTGACCTCGAGGTGGACA  
CCACCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCGGAGCCCAGAGGGAAGCC  
GCAAGAACCCCGCCCGCACCTGCCGTGACCTCAAGATGTGCCACTCTGACTGGAAGA  
GTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCT  
TCTGCAACATGGAGACTGGTGAGACCTGCGTGTACCCCACTCAGCCCAGTGTGGCCC  
AGAAGAACTGGTACATCAGCAAGAACCCCAAGGACAAGAGGCATGTCTGGTTCGGCG  
AGAGCATGACCGATGGATTCCAGTTCGAGTATGGCGGCCAGGGCTCCGACCCTGCCG  
ATGTGGCCATCCAGCTGACCTTCCTGCGCCTGATGTCCACCGAGGCCTCCCAGAACA  
TCACCTACCACTGCAAGAACAGCGTGGCCTACATGGACCAGCAGACTGGCAACCTCA  
AGAAGGCCCTGCTCCTCAAGGGCTCCAACGAGATCGAGATCCGCGCCGAGGGCAACA  
GCCGCTTACCTACAGCGTCACTGTGATGGCTGCACGAGTCACACCGGAGCCTGGG

GCAAGACAGTGATTGAATACAAAACCACCAAGTCCTCCCGCCTGCCCATCATCGATG  
TGGCCCCCTTGGACGTTGGTGCCCCAGACCAGGAATTCGGCTTCGACGTTGGCCCTG  
TCTGCTTCCTG**TAA**ACTCCCTCCATCTAGA

Xba I

&lt;210&gt; 14

&lt;211&gt; 699

&lt;212&gt; PROTEIN

&lt;213&gt; HOMO SAPIENS

&lt;400&gt; 14

1 MNRGVPPFRHL LLVLQLALLP AATQGKKVVL GKKGDTVELT CTASQKKSQ FHWKNSNQIK 60  
61 ILGNQGSFLT KGPSKLNDRA DSRRLWDQG NFPLIKNLK IEDSDTYICE VEDQKEEVQL 120  
121 LVFGLTANS DTHLLQGQSLT LTLESPPGSS PSVQCRSPRG KNIQGGKTLVSQLELQDSG 180  
181 TWTCTVLQNQ KKVEFKIDIV VLAFQKASSI VYKKEGEQVE FSFPLAFTVE KLTGSGELWW 240  
241 QAERASSSKS WITFDLKNKE VSVKRVTDQD KLQMGKKLPL HLTLPQALPQ YAGSGNLTLA 300  
301 LEAKTGKHLQ EVNLVVMRAT QLQKNLTCEV WGPTSPKLML SLKLENKEAK VSKREKAVVW 360  
361 LNPEAGMWQC LLSDSGQVLL ESNIKVLPRS NGLPGPIGPP GPRGRTGDAG PVGPPGPPGP 420  
421 PGPPGPPSAG FDFSFLPQPP QEKAHDGGRY YRADDANVVR DRDLEVDITL KSLSQIQENI 480  
481 RSPEGSRKNP ARTCRDLKMC HSDWKSGEYW IDPNQGCNLD AIKVFNCMET GETCVYPTQP 540  
541 SVAQKNWYIS KNPDKRHWV FGESMTDGFQ FEYGGQGS DPADVAIQLTFL RLMSTEASQN 600  
601 ITHCKNSVA YMDQQTGNLK KALLKGSNE IEIRAEGNSR FTYSVTVDGC TSHTGAWGKT 660  
661 VIEYKTTKSS RLPIDVAPL DVGAPDQEFQ FDVGPVCFL

&lt;210&gt; 15

&lt;211&gt; 1947

&lt;212&gt; cDNA

&lt;213&gt; HOMO SAPIENS

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 24 1931

&lt;400&gt; 15

Hind III

AAGCTTCCCTCGGCAAGGCCACA**AATGA**ACCGGGGAGTCCCTTTTAGGCACTTGCTTC  
TGGTGCTGCAACTGGCGCTCCTCCAGCAGCCACTCAGGGAAAGAAAGTGGTGCTGG  
GCAAAAAAGGGGATACAGTGGAAGTACCTGTACAGCTTCCCAGAAGAAGAGCATAC  
AATTCCACTGGAAAACTCCAACCAGATAAAGATTCTGGGAAATCAGGGCTCCTTCT  
TAACTAAAGGTCCATCCAAGCTGAATGATCGCGCTGACTCAAGAAGAAGCCTTTGGG  
ACCAAGGAAACTTTCCCCTGATCATCAAGAATCTTAAGATAGAAGACTCAGATACTT  
ACATCTGTGAAGTGGAGGACCAGAAGGAGGAGGTGCAATTGCTAGTGTTGGATTGA  
CTGCCAACTCTGACACCCACCTGCTTCAGGGGCAGAGCCTGACCCTGACCTTGGAGA  
GCCCCCTGGTAGTAGCCCTCAGTGCAATGTAGGAGTCCAAGGGGTAAAAACATAC  
AGGGGGGGAAGACCCTCTCCGTGTCTCAGCTGGAGCTCCAGGATAGTGGCACCTGGA  
CATGCACTGTCTTGCAGAACCAGAAGAAGGTGGAGTTCAAAATAGACATCGTGGTG  
TAGCTTTCCAGAAGGCCTCCAGCATAGTCTATAAGAAAGAGGGGGAACAGGTGGAGT  
TCTCCTTCCCACTCGCCTTTACAGTTGAAAAGCTGACGGGCAGTGGCGAGCTGTGGT  
GGCAGGCGGAGAGGGCTTCCCTCCTCCAAGTCTTGATCACCTTTGACCTGAAGAACA  
AGGAAGTGTCTGTAAAACGGGTTACCCAGGACCCTAAGCTCCAGATGGGCAAGAAGC  
TCCCGCTCCACCTCACCTGCCCCAGGCCTTGCCCTCAGTATGCTGGCTCTGGAAACC  
TCACCTGGCCCTTGAAGCGAAAACAGGAAAGTTGCATCAGGAAGTGAACCTGGTGG  
TGATGAGAGCCACTCAGCTCCAGAAAAATTTGACCTGTGAGGTGTGGGGACCCACCT  
CCCCTAAGCTGATGCTGAGCTTGAACTGGAGAACAAGGAGGCAAAGGTCTCGAAGC

GGGAGAAGGCGGTGTGGGTGCTGAACCCTGAGGCGGGGATGTGGCAGTGTCTGCTGA  
 GTGACTCGGGACAGGTCTGCTGGAATCCAACATCAAGGTTCTGCCCAGATCTGATG  
 CCAATGTGGTTTCGTGACCGTGACCTCGAGGTGGACACCACCCTCAAGAGCCTGAGCC  
 AGCAGATCGAGAACATCCGGAGCCCAGAGGGAAGCCGCAAGAACCCCGCCCGCACCT  
 GCCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACC  
 CCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGTG  
 AGACCTGCGTGTACCCCACTCAGCCCAGTGTGGCCGAGAAGAACTGGTACATCAGCA  
 AGAACCCCAAGGACAAGAGGCATGTCTGGTTTCGGCGAGAGCATGACCGATGGATTCC  
 AGTTCGAGTATGGCGGCCAGGGCTCCGACCCTGCCGATGTGGCCATCCAGCTGACCT  
 TCCTGCGCCTGATGTCCACCGAGGCCTCCCAGAACATCACCTACCACTGCAAGAACA  
 GCGTGCCCTACATGGACCAGCAGACTGGCAACCTCAAGAAGGCCCTGCTCCTCAAGG  
 GCTCCAACGAGATCGAGATCCGCGCCGAGGGCAACAGCCGCTTACCTACAGCGTCA  
 CTGTGCGATGGCTGCACGAGTCACACCGGAGCCTGGGGCAAGACAGTGATTGAATACA  
 AAACCACCAAGTCCTCCCGCCTGCCCATCATCGATGTGGCCCCCTTGGACGTTGGTG  
 CCCCAGACCAGGAATTCGGCTTCGACGTTGGCCCTGTCTGCTTCCTGTAAACTCCCT  
 CCATCTAGA

Xba I

<210> 16

<211> 635

<212> PROTEIN

<213> HOMO SAPIENS

<400> 16

1 MNRGVPPFRHL LLVLQLALLP AATQGKKVVL GKKGDTVELT CTASQKKSQ FHWKNSNQIK 60  
 61 ILGNQGSFLT KGPSKLNDRS DSRRLWDQG NFPLIKNLK IEDSDTYICE VEDQKEEVQL 120  
 121 LVFGLTANS DTHLLQGQSLT LTLESPPGSS PSVQCRSPRG KNIQGGKTLS VSQLELQDSG 180  
 181 TWTCTVLQNN KKVEFKIDIV VLAFOKASSI VYKKEGEQVE FSFPLAFTVE KLTGSGELWW 240  
 241 QAERASSSKS WITFDLKNKE VSVKRVTDQ KLMGMKKLPL HLTLPQALPQ YAGSGNLTLA 300  
 301 LEAKTGKHLQ EVNLVVMRAT QLQKNLTCEV WGPTSPKLML SLKLENKEAK VSKREKAVVW 360  
 361 LNPEAGMWQC LLSDSGQVLL ESNIKVLPRS DANVVRDRDL EVDITLKSLS QQIENIRSPE 420  
 421 GSRKNPARTC RDLKMCHSDW KSGEYWIDPN QGCNLDIAKV FCNMETGETC VYPTQPSVAQ 480  
 481 KNWYISKNP KDKRHVWFGES MTDGFQFEYG GQGSDDADVA IQLTFLRLMS TEASQNITYH 540  
 541 CKNSVAYMDQ QTGNLKKALL LKGSNEIEIR AEGNSRFTYS VTVDGCTSHT GAWGKTVIEY 600  
 601 KTKSSRLPI IDVAPLDVGA PDQEFQFDVG PVCFL

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**